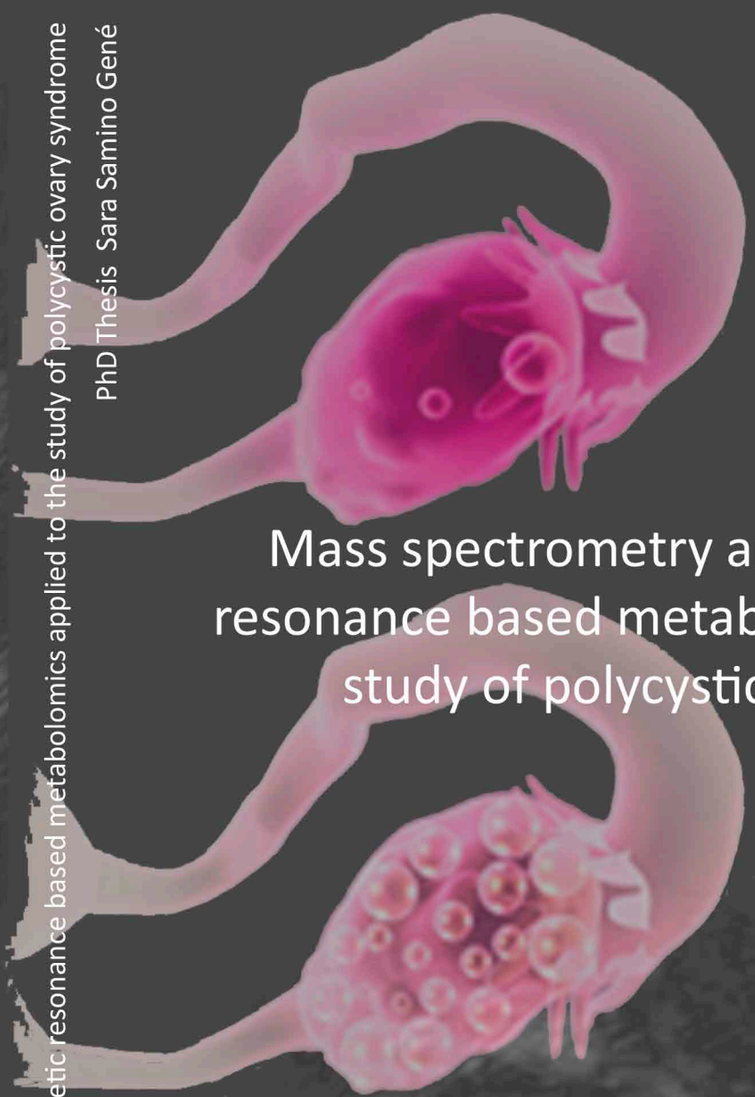
MASS SPECTROMETRY AND NUCLEAR MAGNETIC RESONANCE BASED METABOLOMICS APPLIED TO THE STUDY OF POLYCYSTIC OVARY SYNDROME

Sara Samino Gené

Dipòsit Legal: T.63-2014

UNIVERSITAT ROVIRA I VIRGILI
MASS SPECTROMETRY AND NUCLEAR MAGNETIC RESONANCE BASED METABOLOMICS APPLIED TO THE STUDY OF POLYCYSTIC OVARY SYNDROME
Sara Samino Gené
Dipòsit Legal: T.63-2014

Mass spectrometry and nuclear magnetic resonance based metabolomics applied to the study of polycystic ovary syndrome
PhD Thesis Sara Samino Gené



Mass spectrometry and nuclear magnetic resonance based metabolomics applied to the study of polycystic ovary syndrome

PhD Thesis
Sara Samino Gené

UNIVERSITAT ROVIRA I VIRGILI

Sara Samino Gené

Mass spectrometry and nuclear magnetic resonance based metabolomics applied to the study of polycystic ovary syndrome

Ph. Doctoral Thesis

Directed by Dr. Oscar Yanes Torrado

Departament de Bioquímica i Biotecnologia

UNIVERSITAT ROVIRA I VIRGILI

Tarragona
2013

UNIVERSITAT ROVIRA I VIRGILI

Departamento de Bioquímica i Biotecnología
c/ Marcel·lí Domingo s/n
Campus Sant Pere Sescelades
43007 Tarragona
Teléfono: 977 559 521
Fax: 977 558 232

HAGO CONSTAR que el presente trabajo, titulado “Mass Spectrometry and Nuclear Magnetic Resonance based metabolomics applied to the study of Polycystic Ovary Syndrome” que presenta Sara Samino Gené para la obtención del título de Doctor, ha sido realizado bajo mi dirección en el Departamento de Bioquímica y Biotecnología de esta Universidad y que cumple los requisitos para poder optar a la Mención Internacional.

Tarragona, 17 de Julio de 2013

El director de la tesis doctoral

Oscar Yanes Torrado

"Live as if you were to die tomorrow,

learn as if you were to live forever"

Mahatma Gandi

Agradecimientos

Empezar diciendo que me gustaría agradecer a todo el mundo que ha participado de manera directa o indirecta en la elaboración de esta tesis. Disculpad aquellas personas que han participado también y no son nombradas. Gracias también a vosotros.

Agradecer a las dos personas que me brindaron la oportunidad de realizar esta tesis doctoral, Xavier Correig y Cinta Bladé. Ambos han hecho de esta tesis una tesis muy especial. Agradecerles la oportunidad que me ofrecieron al adentrarme en el mundo de la investigación.

Gracias también a vosotros, los que habéis compartido conmigo estos cuatro años. Tres personas para mi muy especiales, tanto laboralmente como personalmente hablando, ya que sin ellos, esta tesis no hubiese sido posible. Porque me considero una persona muy afortunada por haberles tenido cerca siempre. Gracias, gracias de corazón:

A Miguel, gracias, por tu paciencia y tu conocimiento, siempre tienes la respuesta adecuada a cualquier pregunta!!

A Toni, gracias por todo lo que me has enseñado. Pero, gracias también por tu apoyo incondicional, por tu sonrisa, tu alegría y por darme paz cuando hay tormenta. Ets molt epecial.

Y, a Mariona ¿¡qué decir de ti!? Algunos dicen que eres curiosa, yo te digo que eres emocional, mágica, alocada... durante estos años hemos guardado documentos en *penchys*, hemos hecho *escripis* en *matlapy* y hemos sido *fashionis*! Esa eres tú! Gracias, porque, aunque posiblemente

no eres consciente, me has enseñado tantas cosas... tanto buenas... como malas!!

También agradecerles a Núria y Roger por ser grandes compañeros de viaje, a Josep, Xavi, Rubén, Didac, y como no, a la recién incorporada Míriam, porque aunque ahora te parece que todo es muy difícil, lo sacarás!

También quiero agradecer a mi profesor Jaume Solé, porque me acompañó en mi etapa adolescente, me guió hasta el mundo universitario y me enseñó a luchar por aquello que quería.

Y, como no! Al director, al meu cap! Oscar, por haber apostado por mi cuando el camino estaba a medio recorrer. Por tu paciencia y sencillez, por no despeinarte, por enseñarme a hacer ciencia de una manera tan especial, porque sin ti, esta tesis nunca hubiese sido así!

En el ámbito personal, quiero agradercer primero, a mi familia: en especial a mis padres, por la educación que me disteis y los valores que me enseñasteis, por estar siempre a mi lado apoyándome y confiando en mi. A mi abuela, mi hermana y mi cuñado. Gracias por haber estado siempre a mi lado, gracias por vuestro apoyo. Y gracias a mi sobrina, Ainara, por su inocencia y sus ganas de aprender.

También agradecer a dos personas muy especiales que desafortunadamente no pudieron seguir el camino hasta el final, a los cuales llevo en mi corazón y se que estarían muy orgullosos de este trabajo. Por ellos, por mi tío y por mi abuelo.

Además quiero agradecer a Joan que me animó a emprender este largo camino. Por confiar en mi y apoyarme siempre en mis decisiones.

Además agradecer a mis amigas. A Meri, por estar siempre a mi lado, por escuchar, entender y no juzgar. A Zoraida, porque ha aportado su granito de arena a este libro, mil gracias, pero además, por estar siempre ahí. También a Yasmina, Ari, Esther y Cris, por vuestro ánimo y apoyo, y por que algunas de vosotras también sabéis lo que es esto ;-) Muchas gracias a todas, de todo corazón.

También agradecerle a Víctor, por su apoyo y su confianza, gracias por mostrarme que el camino es sencillo.

Todos y cada uno de vosotros habéis hecho de esta tesis una realidad.

Gracias.

TABLE OF CONTENTS

LIST OF ABBREVIATION

1D	1-dimension
2D	2-dimensions
2D-JRES	2-dimensional J-resolved spectrum
9-HODE	9-hydroxyoctadecadienoic acid
13-HODE	13-hydroxyoctadecadienoic acid
ACN	Acetonitrile
ANOVA	Analysis of variance
APCI	Atmospheric pressure chemical ionization
Apo-AI	Apolipoprotein AI
APPI	Atmospheric pressure photo ionization
ASRM	American society for reproductive medicine
BMI	Body mass index
CE	Capillary Electrophoresis
CHCA	4-hydroxycinnamic acid
$CHCl_3$	Chloroform
CHD	Coronary heart disease
CI	Chemical Ionization
CoA	Coenzyme A
COSY	Correlation spectroscopy
CPMG	1D Carr-Purcell-Meiboom-Gill
CV	Coeficient of variation
CDV	Cardiovascular disease
D_2O	Deuterated water
Da	Dalton
DHB	2,5-dihydroxybenzoic acid
DNA	Deoxyribonucleic Acid

EI	Electron Ionization
ERETIC	Electronic Reference to access In vivo Concentrations
ESHRE	European society for human reproduction and embriology
ESI	Electrospray Ionization
Flut	Flutamide
FT	Fourier Transform
GC	Gas Chromatography
GSH	Glutathione
GSSG	Oxidized glutathione
H_2O	Water
HDL	High density lipoprotein
HILIC	Hydrophylic Interaction Liquid Chromatography
^{1}H-NMR	Proton Nuclear Magnetic Resonance
HPLC	High Performance Liquid Chromatography
HMDB	Human Metabolome Database
IMT	Intima media thickness
IPA	Ingenuity Pathway Analysis
KO	Knock-out
LC	Liquid Chromatography
LCAT	Lecithin cholesterol acyltransferase
LDL	Low density lipoprotein
MALDI	Matrix-Assisted Laser Desorption/Ionization
Met	Metformin
MetOH	Methanol

MetOx	Methionine oxidation
MS	Mass Spectrometry
MS/MS	Tandem mass spectrometry
m/z	mass-to-charge ratio
NIH	National institute of health
NIST	National Institute of Standards and Technology
NOESY-presat	Nuclear Overhauser effect spectroscopy-presaturation
NMR	Nuclear Magnetic Resonance
PCOS	Polycystic Ovary Syndrome
PCA	Principal Component Analysis
Pio	Pioglitazone
QqQ	Triple quadrupole
RNA	Ribonucleic Acid
ROS	Reactive oxygen species
RP	Reverse phase
RT	Room temperature
SHGB	Sex hormone binding-globulin
SOD	Superoxide dismutase
TOCSY	Total Correlation Spectroscopy
TOF	Time of Flight
TSP –d4	3-trimethylsilylpropionic acid –deuterium 4
UHPLC	Ultra-high performance liquid chromatography
VLDL	Very low density lipoprotein
WT	Wild Type

OBJECTIVES

The constant growth of metabolomics over the last 5 years is an undeniable fact that indicates the significant contribution of this discipline to different fields of research.

In this context, the three objectives of this thesis have been:

1. Mastering of the main analytical platforms used in metabolomics: nuclear magnetic resonance (NMR), liquid chromatography (LC), and gas chromatography (GC) coupled to mass spectrometry (MS).

2. Developing an untargeted metabolomic workflow, involving novel aspects of sample preparation, and data processing for metabolite identification.

The achievement of these two analytical and methodological objectives has led to the publication of two peer-review articles, which consitute Work 1 and 2 of Chapter 2.1:

In Work 1: I aimed to optimize metabolite extraction conditions for NMR analysis, followed by LC-ESI-MS by using the same sample extract with no need for solvent exchange or further pretreatment.

In Work 2: I aimed to investigate the impact of different aspects of univariate statistical analysis on untargeted LC-MS based metabolomic experiments.

3. Implementing our untargeted metabolomic workflow to the study of human patients with Polycystic Ovary Syndrome (PCOS) and their response to drug treatment.

These clinical studies led to the publication of two peer-review articles, and a third work submitted for publication which consitute Works 3-5 of Chapter 2.2:

Objectives

In Work 3: I aimed to implement a GC-MS untargeted metabolomic approach to provide new insights on the impact that obesity exerts on the metabolic derangements associated with PCOS.

In Work 4: I aimed to implement a multiplatform approach based on NMR and LC-MS to provide new insights in PCOS disease in a cohort of young lean PCOS patients.

In Work 5: I aimed to implement a multiplatform approach based on NMR, GC-MS and LC-MS to provide new insights on the action of drug polytherapy to PCOS disorder.

CHAPTER 1. INTRODUCTION

1.1 METABOLOMICS

Metabolomics is an emerging field in the "omics" scene that focuses on the comprehensive characterization of small molecules in biological matrices. These small molecules, or metabolites, are intermediates and products of metabolism. Metabolites are involved in a diversity of cellular functions, including cell energetics, inflammation, signalling, as well as building blocks of structural biopolymers such as proteins and DNA [1]. Endogenous metabolites are often altered in response to environmental factors, disease, nutrition and other aspects in an attempt to maintain cellular homeostasis in the organism. Metabolomics is able to cope with such disturbances providing a comprehensive view of biochemical reactions and cellular phenotypes. The total collection of metabolites in a cell, tissue or organism is defined as the "metabolome" [1]. A full understanding of the metabolome is essential for obtaining a comprehensive view of the functioning of cells, tissues and organisms.

The word metabolomics was coined by analogy with other omic approaches, and has emerged as a complementary field to genomics, transcriptomics and proteomics, which are involved in the global study of DNA, mRNA, and proteins, respectively. The metabolome is considered the end-point of the 'omics' cascade [2]. Genes and proteins are subject to epigenetic regulation and post-translational modifications, respectively, while metabolites provide an instantaneous *snapshot* of the physiology of the biological system under study. Changes in the metabolome are the ultimate answer of an organism to genetic alterations, disease, or environmental influences. These changes, therefore, reflect more accurately the phenotype of a biological system.

In this context, recent innovations in instrumentation in MS and NMR, bioinformatic tools, and software enable the comprehensive analysis of cellular metabolites without bias [2].

1.2 TARGETED vs. UNTARGETED METABOLOMICS

The first step in performing metabolomics is to determine the number of metabolites to be measured. Ultimately, the number and chemical composition of metabolites studied is a defining attribute of any metabolomic experiment that shapes experimental design with respect to sample preparation and choice of instrumentation. This important attribute of metabolomic experiments has led the field to refer to either targeted or untargeted approaches.

Targeted metabolomics: This approach is typically hypothesis-driven, and refers to a method in which a small and well-defined set of known metabolites are measured and quantified [3]. Generally, this set of metabolites is focused in one or more related pathways [4]. The targeted analysis of metabolites has been done for many decades now, far before the term metabolomics was coined. A large number of methods are available for many metabolites such as sugars [4], amino acids [5], steroids [6] or fatty acids [7], among others. The benefit of such targeted methods is that minimal effort and resources are required to profile these specific metabolites over a large number of samples.

Untargeted metabolomics: Untargeted methods are global in scope and have the aim to simultaneously measure as many metabolites as possible from biological samples without bias [2]. This metabolomic approach is typically hypothesis generating and provides a global comparative overview of metabolites abundances between two or more sample groups (i.e., experimental conditions), for example, healthy vs. disease or, WT (wild type) vs. KO (knock-out). However, in contrast to targeted metabolomics, the main bottleneck of an untargeted approach is the identification of metabolites on the basis of MS and/or NMR peaks

from exceedingly complex datasets [2]. Despite this challenge, this approach has the potential to involve previously unrecognized metabolites in seemingly known human pathological conditions. As a result, most of the methodological work and clinical studies presented in this thesis will focus on an untargeted metabolomic approach.

1.3 THE UNTARGETED METABOLOMICS WORKFLOW

Although untargeted metabolomic experiments are often hypothesis generating rather than hypothesis driven, it is important to carefully construct an experimental design that maximizes the number of metabolites detected and their quantitative reproducibility.

In contrast to genomics and proteomics that deal with combinations of 4 bases and 20 amino acids, respectively, metabolomics must cope with the great chemical and physical complexity of metabolites, ranging from very polar to very hydrophobic compounds and a wide mass window, typically between 50 Da and 1500 Da [1]. In addition, some of them are present at low pico-molar concentrations such as hormones and neurotransmitters, and other at milli-molar concentrations such as glucose or some amino acids.

Therefore, in order to obtain reliable results from untargeted metabolomic studies, numerous factors must be carefully considered. These factors include: sample preparation, implementation of appropriate MS and NMR analytical tools for sample analysis, data analysis, and last but not least, biological interpretation [8] (Figure 1). Hence, a well-defined workflow characterizes untargeted metabolomics. The following sections will consider specific issues related to this workflow.

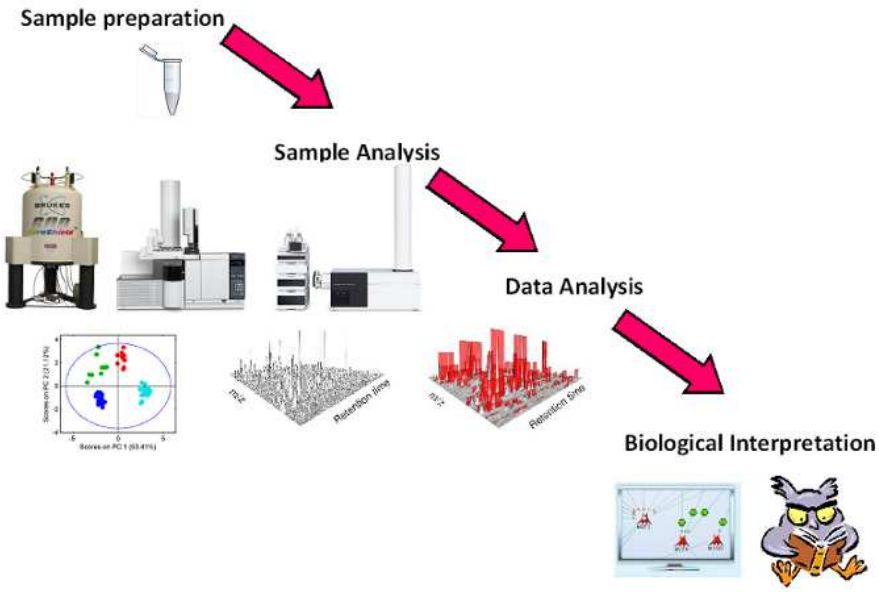


Figure 1. Untargeted metabolomics workflow

1.3.1 Sample preparation

In metabolomics, metabolites are typically isolated from complex biological matrices, including serum, plasma, tissue homogenates, saliva, urine, or cell pellets.

Sample quality preparation is a key factor that determines the success of any analytical procedure [9]. Ideally, metabolite extraction aims to (i) efficiently release metabolites from sample, (ii) remove interferences that hinder a proper analysis (e.g. proteins and salts) and (iii) make the extract compatible with the analytical technique [10]. The overall purpose of the sample preparation procedure is to reproducibly transform the sample into a format that will be compatible with the analytical technique chosen whereas the original metabolite composition is maintained. The optimal method for untargeted

metabolomics should (a) extract the largest number of metabolites, (b) be nonselective and not exclude molecules with particular physical or chemical properties, and (c) be nondestructive or modify metabolites through chemical or physical means.

Due to the large chemical diversity of metabolites, however, no universal isolation method for detecting all metabolites is available. Generally, organic solvents are used for sample homogenization, protein precipitation and solubilisation of metabolites [9,10]. The solubility of a metabolite fundamentally depends on the organic solvent, temperature and pH used. Polar solvents (e.g methanol [11], methanol-water [12] mixtures or ethanol [9]) are normally used in metabolomics expermients. But, non-polar solvents such as chloroform or hexane can be used to extract lipophilic components. Metabolites can be also extracted at extreme pH (such as perchloric acid [13] or meta-phosphoric acid [9]) with the aim to extract acid-stable compounds.

Therefore, the method used will ultimately determine the classes of metabolites detected by MS or NMR. The wide range of metabolites with different physicochemical properties usually demands application of complementary sample preparation protocols.

1.3.2 Sample Analysis

Due to the large chemical complexity of the metabolome, it is not possible to analyse the entire range of metabolites using a single analytical platform. Consequently, multiple analytical platforms are needed to increase the coverage of the metabolome [8]. NMR and MS are the most widely used analytical platforms in metabolomics [14]. MS

is normally coupled to chromatographic techniques to extend the range of compound classes detected. In short, LC, GC and Capillary Electrophoresis (CE) coupled to MS, together with NMR have emerged as the analytical platforms used in metabolomics experiments [15,16,17].

1.3.2.1 Mass Spectrometry

MS is an analytical technique that can provide both qualitative (structure) and quantitative (molecular mass and concentration) information of metabolites after their conversion to ions. Metabolites are first introduced into the ionisation source of the mass spectrometer, where they are ionised to acquire positive or negative charges. The ionized molecules then travel through the mass analyser and arrive at the detector according to their mass-to-charge (m/z) ratio. Mass analysers measure the accurate mass of each metabolite, thus is possible to assign an elementary composition to each mass. In addition to this qualitative description, mass spectrometry offers a measure for quantity by counting the ion abundance of each mass. This is done after the ions make contact with the detector, where useable signals are generated and recorded by a computer system. The computer displays the signals graphically as a mass spectrum showing the relative abundance of the signals according to their m/z ratio (Figure 2).

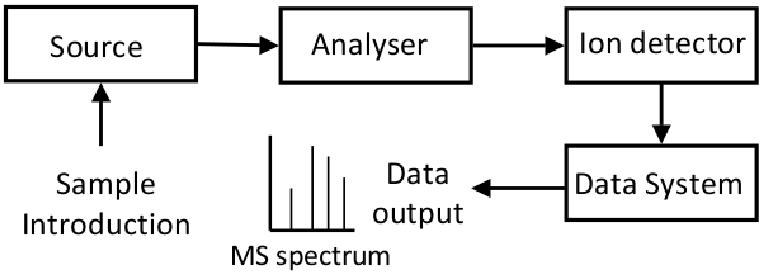


Figure 1. General workflow of a mass spectrometer

The sample can be inserted directly into the ionisation source, or can undergo some type of chromatography before entering the ionisation source. In metabolomics experiments, this later method is generally used [18]. Therefore, sample introduction usually involves a mass spectrometer coupled directly to a high pressure liquid chromatography (HPLC) [19], ultra high pressure liquid chromatograpy (UHPLC) [20], GC [16] or CE [15] (Figure 3). Hence the sample is separated into a series of components that then enter the mass spectrometer sequentially for individual analysis. In this thesis, the separation methods used have been GC and LC, and they are detailed in the following section.

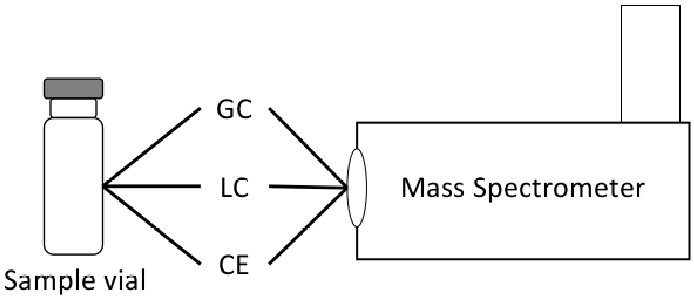


Figure 2. Chromatography coupled to Mass Spectrometry

Gas Chromtography-Mass spectrometry

GC-MS has been applied to a wide range of investigations of the human metabolome [21,22].

In GC-MS, sample extract is introduced into a heated injector (200-250°C), where rapid vaporization and mixing with the carrier gas (usually helium) occur, followed by chromatographic separation of metabolites on the GC column and subsequent MS detection.

Sample can be injected in either split or splitless mode. In a splitless system, the advantage is that larger amount of sample can be introduced into the column. However, a split system is preferred when the detector is sensitive to trace amounts of analyte and there is concern about sample overloading in the column. In metabolomics studies, split mode is generally preferred because metabolites are present in a wide range of concentrations. Then, small molecules are separated in GC columns based on their volatility, or ease with which they evaporate into a gas, before they enter into the mass spectrometromer. In general, small molecules travel more quickly than larger molecules. The retention of such metabolites is based on the partitioning between the mobile phase, consisting of a carrier gas, and a stationary phase consisting of a liquid residing on the inside wall of the capillary. The stationary phase may be of many different types but, in metabolomics approaches, usually is a non-polar phase with high intertness. The capillary column is held in an oven that can be ramped continuously or in steps to achieve desired separation. As the temperature increases, those compounds that have low boiling points elute from the column sooner than those that have higher boiling points. Typical column oven temperatures range from 40

to 325ºC [23]. The rate at which a sample passes throug the column is directly proportional to the temperature of the column.

The gas chromatograph is typically coupled to a mass analyser via an electron ionization (EI) source [21]. The EI source is now widely implemented in metabolomics, but other ion sources can also be used, including chemical ionisation (CI) [24]. In EI, analytes are fragmented into ions by a stream of electrons with a kinetic energy of 70 eV. All these ions are separated according to their mass-to-charge ratio (m/z) and ions are detected in proportion to their abundance. The fragmentation of molecules is highly dependent on the chemical structure. Each metabolite presents a particular fragmentation pattern called mass spectrum that can be compared with GC-MS libraries such as the National Institute for Standard Technology (NIST).

. There are two main mass analysers that can be coupled to GC and used in untargeted metabolomics. These are single quadrupoles [25], which provide nominal-mass information, or TOF systems [21,26], which depending on the vendor provide nominal-mass or high-mass accuracy. TOF analyzers provide faster scan-rates than single quadrupoles, and recently, GC-quadrupole-TOF instruments have been implemented that provide MS/MS capabilities [27].

Triple quadrupole (QqQ) mass analysers are also used for metabolomics experiments coupled to GC, but generally, QqQ are only used for targeted metabolomic analyses [28]. QqQ mass analyzers enable higher level of molecular specificity and selectivity. In this thesis, however, GC-MS is performed using single quadrupole and QTOF mass analyzers.

The advantages of GC analysis are its high separation efficiency and robustness [29]. The main drawback is that it can only be used to separate volatile analytes. The vast majority of metabolites are non-volatile, imposing the necessity to derivatise metabolites. However, many metabolites can not become volatile after derivatization either, particularly large metabolites that form strong intermolecular interactions. Another drawback is the degradation of thermolabile metabolites. Therefore, comprehensive techniques such as LC-MS or NMR have to be used to cover a greater diversity of metabolites.

High-Performance Liquid Chromatography/mass spectrometry

Like GC, LC separation is based on partitioning, but in this case between a liquid flow (mobile phase) and a stationary phase. HPLC relies on pumps to pass a pressurized liquid and a sample mixture through a column filled with a sorbent, leading to the separation of the sample components (i.e., metabolites). The active component of the column, the sorbent or stationary phase, is typically a granular material made of solid particles (e.g. silica, polymers, etc.), 2-50 micrometers in size. Metabolites are separated from each other due to their different degrees of interaction with the sorbent particles. The pressurized liquid is typically a mixture of solvents (e.g. water, acetonitrile and/or methanol) and is referred to as a "mobile phase". Its composition and temperature plays a major role in the separation process by influencing the interactions taking place between metabolites and sorbent. These interactions are physical in nature, such as hydrophobic, hydrophilic and ionic, most often a combination thereof. Hence, metabolites do not have

to be volatile and therefore do not requiere derivatisation prior to analysis. Common stationary phases in metabolomics are reverse phase (RP) C-18 and C8, and hydrophilic interaction liquid chromatography (HILIC) [30]. C-18 chromatography is the standard tool for the separation of medium polar and non-polar metabolites. However, very polar metabolites are not retained in classical C-18 stationary phases and elute within the void volume. The alternative is to use HILIC chromatography [31]. HILIC provides a method to more reliably analyze strongly polar and ionic metabolites. Aside from better retention of polar metabolites, another advantage often reported for HILIC methods is an improvement in MS signal intensity due to the high organic-content mobile phases that enable more efficient electrospray droplet desolvation [32]. It is important to note, however, that this increased MS sensitivity may not necessariy result in the observation of more metabolic features depending on the samples type analyzed. Just as highly polar compounds elute in the void volume of C-18, highly hydrophobic metabolites elute in the void volume of HILIC and ion supression can therefore cause the number of lipid features to decrease substantially. In this perspective, C-18 and HILIC methods are complementary with respect to metabolite coverage and best used in parallel to accomplish global analysis [33].

As with GC, after liquid chromatographic separation, metabolites need to be ionized prior to analysis in the mass spectrometer. In the ionization source, analytes are transferred from liquid phase into gas phase. This is generally done by spraying the sample through a metal capillary by a turbulent air stream. Based on this principle three major ionization methods are available: atmospheric pressure photo ionization

(APPI) [34], atmospheric pressure chemical ionization (APCI) [35], and electro spray ionization (ESI) [30]. In this thesis, ESI has been the chosen ionization. ESI is a "soft" ionization technique, provided that there is very little fragmentation of analyte structures. In order to obtain a broad coverage of the metabolome, ionization must be performed in positive and negative mode [36]. ESI is the most efficient interface for generating $[M+H]^+/[M-H]^-$ (positive/negative) ions but it also produces adducts such as [M+Na]+ or [M+K]+ ions by thermal ion attachments [30].

From the ionization source ions are guided into the mass analyser. In untargeted metabolomics analyses the ESI source is commonly coupled to high resolution mass analyzers such as Time of Flight (TOF) [37], Orbitrap [38], Quadrupole Time of Flight (QTOF) [9] and Fourier Transform Ion Cyclotron Resonance (FT-ICR) [30]. The mass analyser chosen throughout this thesis is the QTOF (Figure 4). QTOF MS in combination with LC is one of the most used instruments for metabolomics studies [20,39]. Even though it offers moderate mass resolution and mass accuracy when compared with Orbitrap or FT-ICR, QTOF may compensate with very fast scanning rates [30].

QTOF technology provides both a quadrupole and time-of-flight mass analysers with an intermediate collision cell for possible tandem MS fragmentation experiments [40]. A quadrupole mass analyser consists of four metal rods arranged in parallel where those opposite to one another are electrically connected by a radio frequency (RF) voltage supply [40]. The charged molecules enter the quadrupole but only ions within a certain m/z range will survive all the way through the quadrupole and arrive to the collision cell, where they can be

fragmented. The resulting fragment ions reach the TOF for mass accurate measurement and generate a fragmentation spectrum that can be compared with those present in databases. The TOF is a high resolution mass analyzer and functions by applying high voltage pulses to orthogonally accelerated ions into a high vacuum flight tube and a reflectron to reflect them back towards a detector [40]. In untargeted metabolomics studies, however, QTOF is initially used in profile mode. This mode let all ions travel through the quadrupole for accurate measurement by the TOF.

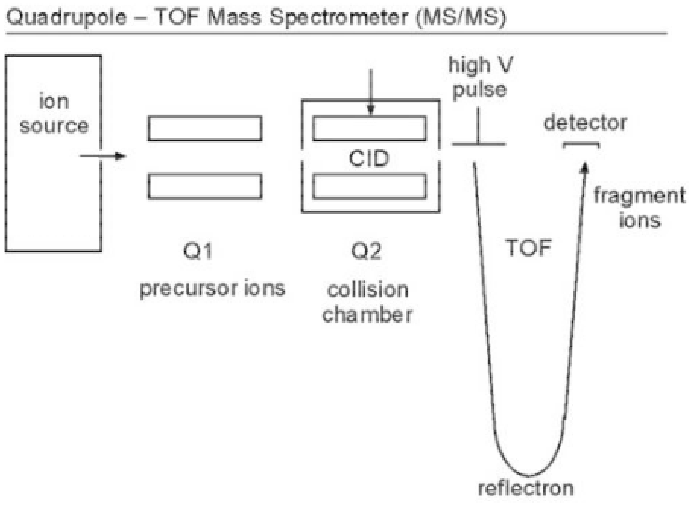


Figure 3.Schematic representation of QTOF mass analyzer.

Alternatively, as with GC-MS, LC can be coupled to a QqQ mass analyzer for targeted analyses [41].

The primary advantage of LC-ESI MS over GC-EI MS is its ability to analyse polar and non-polar thermolabile metabolites [1]. This advantage results in improved coverage of the metabolome. The main

disadvantage of LC-MS with respect to GC-MS is its lower reproducibility and lower robustness.

1.3.2.2 NMR

Along with MS, NMR is the other important analytical platform used in metabolomics [42,43]. NMR is a non-destructive, rapid and highly robust technique that produces highly informative structural and quantitative information. In fact, ^{1}H-NMR spectra can be acquired in a few minutes with minimum sample preparation which usually just entails buffering and/or internal standard addition [44]. The compound used as internal reference in the NMR spectrum is generally the sodium salt of 3-trimethylsilylpropionic acid –d4 (TSP-d4) with deuterated methylene groups, which is used as a reference for concentration and for chemical shift (δ = 0.00 ppm) [44]. Alternatively, a reference signal can be introduced electronically, using an Electronic Reference to access In vivo Concentrations (ERETIC) for quantification purposes [45].

NMR-based metabolomic experiments present to major experimental issues [44]. The first one deals with accurate solvent suppression, and it is generally resolved with a presaturation step. The second experimental issue refers to distinction between small molecular weight metabolites and macromolecules. Macromolecules produce broad resonances due to limited rotational diffusion and short T2 relaxation times, causing difficulties in spectral interpretation. To overcome these problems, 1D nuclear Overhauser effect spectroscopy with presaturation (1D NOESY-presat) [46] and the 1D Carr-Purcell-Meiboom-Gill (CPMG) are two pulse sequences frequently used for

metabolic profiling [47]. 1D NOESY has become the most popular sequence for NMR-based metabolomic analysis [46]. This is mostly due to its efficient water suppression with little calibration and high signal reproducibility. CPMG is a special pulse sequence used to remove broad protein signals, and only small molecules are visible in the spectra [44]. In contrast, LED diffusion provides information on relatively large molecules such as lipoproteins. In this thesis work, I have used 1D LED diffusion NMR experiments for estimating the sizes and the relative proportions of different lipoprotein subclasses followed by a surface fitting algorithm based on Lorentzian functions [52].

Besides 1D ^{1}H NMR, another frequently used sequence in NMR metabolomics studies is homonuclear 2D J-resolved [48]. This sequence increases the identification of biochemical substances by determining the coupling between two different nuclei (e.g., ^{31}P coupled to ^{1}H). Additional 2D NMR spectroscopy methods used in metabolomics are correlation spectroscopy (COSY) [49] and total correlation spectroscopy (TOCSY) [50], which provide spin-spin coupling connectivities [51]. In these methods, magnetization transfer occurs between nuclei of the same type, through J-coupling of nuclei connected by up to a few bonds. In this thesis work, I have used 2D NMR experiments for reliable metabolite identifications.

1.3.2.2 NMR vs. MS

Comparatively, MS and NMR have their own specific advantages and limitations when dealing with metabolomic studies. The main advantage of MS is its sensitivity, as state-of-the-art mass spectrometers can detect

metabolites in the femtomole range [30]. Coupling MS with LC or GC enables the measurement of several hundred individual species within a single analysis [30]. Perhaps the major weakness of MS in metabolomics is absolute quantification. The type of sample preparation used and its molecular environment affect MS signal intensitiesfor every compound. Adding known amounts of internal isotope-labeled standards enables accurate quantification for specific molecules, however, this strategy is impractical for purely discovery-driven metabolomics research, and, for this reason, untargeted metabolomic studies are based on comparing peak area or intensity to establish differences in the relative abundance of specific metabolites between different sample conditions.

This specific weakness of MS is the major strength of NMR. The peak area of a compound in the NMR spectrum is directly related to the concentration of specific nuclei (generally ^{1}H), making absolute quantification of compounds in a complex mixture very precise [53]. Another valuable feature of NMR is its versatility for analyzing metabolites in the liquid state (e.g., serum or urine) or in intact tissues (e.g., liver) without any step of metabolite extraction. Unfortunately, sensitivity, which is the major strength of MS, is the major weakness in NMR. NMR is still order of magnitude less sensitive than MS.

Taken together, the complementary nature of both NMR and MS technologies provides an opportunity to determine the identity of a large number of metabolites, which is the main objective of an untargeted metabolomic experiment [54]. However, sample preparation disparities between NMR and MS may limit the analysis of the same samples by both platforms. In this thesis, this issue has been addressed by

investigating different extraction protocols that may satisfy the requirements for both technologies simultaneously (section 2.1.1).

1.3.3 Data Analysis

In contrast to targeted metabolomics, untargeted metabolomic datasets are exceedingly complex with file sizes on the order of gigabytes per sample for some new high-resolution MS instruments [18]. Manual inspection of the hundreds and thousands of peaks detected by NMR and MS, respectively, is impractical and complicated by experimental drifts in instrumentation. In LC-MS experiments, for example, there are deviations in retention time from sample to sample as a consequence of column degradation, carry over, small fluctuations in room temperature, mobile phase pH, etc. Therefore, these and other challenges may present significant obstacles for interpreting and converting untargeted profiling data into useful biological information. In this thesis work, major progress has been made to establish a robust data analysis workflow for MS-based metabolomics making especial emphasis on statistical analysis to rank relevant metabolite features.

Two possible workflows for data analysis are proposed depending on the analytical workflow (Figure 5):

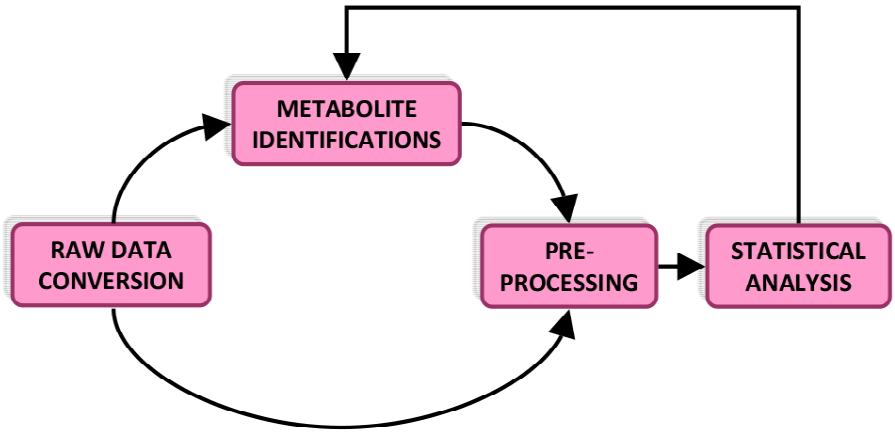


Figure 4. Data analysis workflow for untargeted metabolomics experiments.

(a) **Feature-oriented data proc**essing:

Firstly, raw data is converted into a standard format such as mzData. Vendor or free software (e.g., ProteoWizard [55]) are available for generating mzData.

Several MS-based metabolomic software provide methods for peak picking, non-linear retention time alignment, visualization, and relative quantitation. The XCMS software [56] has been used throughout this thesis, however, there are additional tools such as Mzmine [57] or Metaling [58] that can also provide automatic methods for peak picking, alignment, and relative quantification. XCMS is implemented as a package within the R programming environment. XCMS implements three main steps:

1. *Peak detection*: the algorithm identifies peaks in each of the samples;

2. *Non-linear retention time alignment*: matches peaks with similar retention times across multiple samples, and use the groups of matched peaks for time alignment;

3*. Fill in any missing peaks:* the peaks that have failed to be detected in step 1, are filled in properly from raw data.

As a result, the output matrix of XCMS includes the intensity or area under the peak for all features detected, for every sample. In this process, thousands of so called metabolite features are routinely generated. A feature is defined as a peak corresponding to an individual ion with a unique mass-to-charge ratio and a unique retention time. It is worth emphasizing that one feature is not equivalent to one metabolite. Generally, more than one feature belongs to the same metabolite. For example, due to "hard" ionization of GC-EI MS, the fragmentation of metabolites generates more than one feature for the same compound. In LC-MS data, isotopic distributions, potential adducts or in-source fragmentation generates more than one feature for the same metabolite. This redundancy can be filtered and annotated using several open-source software such as CAMERA [59] and/or Astream [60], or commercially available tools such as Mass Hunter (Agilent Technologies) or Sieve (Thermo Scientific).

Next, the abundance of all these features can be normalized to internal standards (particularly in GC-MS) and statistical analysis will reveal which features are significantly different between sample groups. Finally, in a feature-oriented workflow, the last step is metabolite identification.

- Identification of metabolites by GC-MS: statistically significant features are assigned to metabolites by comparing their mass spectra to those present in libraries such has NIST, Fiehn or Golm Metabolome Database [61].
- Identification of metabolites by LC-MS: to determine the identity of a feature of interest, the accurate mass of the compound is first searched in metabolite databases such as Metlin [62] or the Human Metabolome Database (HMDB) [63]. A database match represents only a putative assignment that must be confirmed by comparing the retention time and MS/MS data of a model compound to that from the feature of interest in the research sample. Currently, MS/MS data for features selected from the profiling results are obtained from additional experiments and matching of MS/MS fragmentation patterns is performed manually by inspection. These additional analyses are time-consuming and represent the rate-limiting step of the untargeted metabolomic workflow.

(b) **Metabolite-oriented data processing**: in this workflow metabolites are identified before any statistical analysis, and it is commonly implemented for GC and NMR-based experimental designs.

- NMR: raw data are typically processed by means of Fourier Transform (FT) and baseline removal [44]. Baseline distortions affect not only the statistical analysis but also the quantification of metabolites. These distortions can be corrected in many different ways but usually an automated baseline correction is applied. Next, all spectra are aligned by

spectral referencing [64]. Finally, two strategies can be followed for metabolite identification: (i) selected peaks in the NMR spectra are integrated and identified using the AMIX software; (ii) NMR resonances are deconvoluted and metabolites identified by manual fitting using Chenomx database and commercial software. The areas under the NMR peaks are normalized either to internal standards (e.g., Chenomx and AMIX) or to an electronic reference (e.g., AMIX) (see section 1.4.2). Finally, statistical analysis is performed.

- GC-MS: metabolite identifications can be done automatically by deconvoluting spectra using vendor software such as MassHunter (Agilent Technologies) or CromaTOF (LECO). Both software use Fiehn Metabolomics and NIST libraries for metabolite identification. Alternatively, there are freely available software packages that can be also used, including BinBase [65], TargetSearch [66] or ADAP [67].

1.3.4 Biological interpretation

Metabolomics provides a tool to measure biochemical activity directly by monitoring the substrates and products transformed during cellular metabolism. Untargeted profiling of these chemical transformations at a global level serves as a phenotypic readout that can be used effectively in clinical diagnostics, to identify therapeutic targets of disease, and to investigate the mechanisms of fundamental biological processes. In this sense, there are numerous computational tools to interpret alterations in metabolite abundances, including pathway-oriented databases such as KEGG [68], Cytoscape [69], Reactome [70], or commercially available software such as Ingenuity Pathways Analysis

(Qiagen) and GeneSpring (Agilent Technologies). Additionally, the HMDB contains chemical, clinical and biochemical information (i.e., MetaboCards) for most endogenous metabolites.

Besides these tools, generally, the most effective approach to put metabolite changes into their appropiate biological context is to search thoroughly in the literature.

1.4 CLINICAL METABOLOMICS: PCOS

Considering its sensitivity, throughput, and minimal sample requirements, untargeted metabolomics has wide applicability across a myriad of biological questions. Despite its relatively recent emergence as a global profiling technology, untargeted metabolomics has already affected our understanding of comprehensive cellular metabolism and been utilized to address a number of biomedical issues [71,72,73]. For example, metabolite profiling has become a powerful approach in clinical studies, aiming at acquiring knowledge on the mechanisms of drug action, or the disease itself, and discovering potential new biomarkers.

Within this context, polycystic ovary syndrome (PCOS) has been studied in this thesis with a three-fold purpose (section 2.2): (i) discover potential new biomarkers, (ii) study drug response, and (iii) provide novel biochemical insight into this clinical syndrome.

PCOS is a complex endocrine disorder characterised by reproductive and metabolic disturbance – polycystic ovaries, anovulation, hyperandrogenism, insulin resistance, hyperinsulinaemia, pancreatic beta cell dysfunction and hyperlipiademia. It is the most common cause worldwide of ovarian infertility and is a familial polygenic condition, linked genetically to both type 2 diabetes and the metabolic syndrome [74,75]. The prevalence of PCOS is between 6% and 10% in developed countries, although this clinical syndrome sits atop a spectrum of disordered polycystic ovarian morphology and metabolic function with an estimated prevalence of 20-30% in developed countries including up to 52% of South Asian immigrant women in Britain [76].

Symptoms

Women with PCOS present a wide range of symptoms. The most common are menstrual irregularities. Most women with PCOS have polycystic ovaries on ultrasound (75-90%) and biochemical or clinical signs of androgen excess (70%). The androgen excess is manifested mainly by hirsutism (i.e. abnormal growth of hair in a person's face and body). PCOS is also one of the most frequently causes of infertility and the most common cause of ovulatory dysfunction[77].

Diagnosis of PCOS

The definition and hence the diagnosis of PCOS remain controversial. To date, two major criteria have been proposed based on two international conferences:

- National Institutes of Health (NIH) criteria (1990) includes: (i) hyperandrogenism and/or hyperandrogenemia, (ii) chronic anovulation, and (iii) exclusion of related disorders such as hyperprolactinemia, thyroid disorders, and congenital adrenal hyperplasia [78]. Conference participants recognized that while polycystic ovaries were often observed, they felt that this finding was suggestive, but not diagnostic of the syndrome.

- European Society for Human Reproduction and Embryology (ESHRE) and the American Society for Reproductive Medicine (ASRM): the proceedings of the conference noted that PCOS could be diagnosed by having two of the following three features: (i) oligo- or anovulation, (ii) clinical and/or biochemical signs of hyperandrogenism, and (iii) polycystic ovaries, after the exclusion of related disorders [79]. These

recommendations did not eliminate the NIH 1990 criteria; rather they expanded the definition of PCOS.

Additional phenotypes now to be considered as PCOS included: (i) women with polycystic ovaries and ovulatory dysfunction but no signs of androgen excess, and (ii) wonem with polycystic ovaries with clinical and/or biochemical evidence of androgen excess, but no signs of ovulatory dysfunction [80].

Treatment options

- Weight loss: Among women with PCOS, 65-75% are obese. Obesity has a significant impact on reproductive outcome. In this case, women should be provided with assistance to lose weight, including psychological support, dietary advice, exercise classes, and, where appropiate, weight-reducing agents or bariatric surgery.

- Hormonal supression: oral contraceptives are the first-line therapy for most of these women. In spite of menstrual cycle regulation, oral contraceptives reduce the biologic activity of androgens. Oral contraceptives contain a combination of estrogen and progesterone.

- Ovulation induction: PCOS is the most common cause of ovulatory dysfunction and one of the most frequently seen indentifiable causes of infertility.

- Metformin: is a biguanide antihyperglycemic agent approved for the treatment of type 2 diabetes mellitus. When it is used in anovulatory women with PCOS, it acts to decrease insulin levels and increase levels of sex hormone-binding globulin (SHGB), which in turn, increases levels of estrogens and decreasesthose of testosterone.

In spite of considerable research efforts [75,81,82], mechanisms leading to PCOS remain unknown, although, like most complex heterogeneous diseases, both environmental and genetic factors are implicated.

CHAPTER 2. RESULTS

2.1 Current Challenges in Untargeted Metabolomics

2.1.1 Assessment of compatibility between extraction methods for NMR- and LC-MS-based metabolomics

SUMMARY

Because of the wide range of chemically and structurally diverse metabolites, efforts to survey the complete metabolome rely on the implementation of multiplatform approaches based on NMR and MS. Sample preparation disparities between NMR and MS, however, may limit the analysis of the same samples by both platforms. Specifically, deuterated solvents used in NMR strategies can complicate LC-MS analysis as a result of potential mass shifts, whereas acidic solutions typically used in LC-MS methods to enhance ionization of metabolites can severely affect reproducibility of NMR measurements. These intrinsically different sample preparation requirements result in the application of different procedures for metabolite extraction, which involve additional sample and unwanted variability. To address this issue, we investigated 12 extraction protocols in liver tissue involving different aqueous/organic solvents and temperatures that may satisfy the requirements for both NMR and LC-MS simultaneously. We found that deuterium exchange did not affect LC-MS results, enabling the measurement of metabolites by NMR and, subsequently, the direct analysis of the same samples by using LC-MS with no need for solvent exchange. Moreover, our results show that the choice of solvents rather than the temperature determined the extraction efficiencies of metabolites, a combination of methanol/chloroform/water and

methanol/water being the extraction methods that best complement NMR and LC-MS analysis for metabolomic studies.

INTRODUCTION

Metabolomics is focused on the profiling and quantification of small, naturally occurring compounds that collectively constitute the so-called metabolome and, as such, serve as direct signatures of biochemical activity in cells.[1-3] In this context, two different technologies have arisen in the field: NMR and MS. As a result of the wide range of chemically and structurally diverse metabolites, however, there is not a unique NMR or MS analytical platform able to reliably measure such a diversity of compounds, especially for untargeted metabolomic studies. Therefore, the implementation of multiplatform approaches has been used to expand coverage of the metabolome.[4] ^{1}H NMR is characterized by a high technical reproducibility, fast analysis, and robust quantification of compounds. In addition, it does not require extensive sample preparation, and the comprehensive databases available facilitate identification of metabolites.[5] The inherent limitation of NMR, however, is its poor sensitivity, which hampers the detection of important metabolites present at low concentrations. In contrast, LC-ESI-MS, the most commonly used analytical combination in MS-based untargeted metabolomics, provides higher sensitivity and molecular specificity.[6] Yet, the lack of many pure standard compounds and comprehensive metabolite MS/MS databases hinders fast and reliable identification of a large number of metabolites. Nevertheless, the distinctive physicochemical principles of NMR and MS make these two technologies

complementary and can potentially expand metabolite coverage in untargeted metabolomic studies. For example, metabolites such as glucose or lactic acid, which are not readily ionized by ESI or not retained using reverse-phase chromatography, respectively, can be detected by using NMR straightforwardly. Such complementarity should gather more information about the levels of metabolites perturbed, thus enabling a wider interrogation of the fundamental biochemical processes involved in the underlying phenotype. The fact, however, is that the vast majority of metabolomic studies have been performed by using a single analytical platform based on either NMR[7–9] or MS[10–13] technology. Few studies have taken full advantage of the complementarity between NMR and MS.[4,14–18] Apart from the need of expertise in MS and NMR and the availability of both technologies, the main reason for using a single platform is the disparity between the initial amount of sample required for NMR and MS analyses and sample preparation (i.e., extraction of metabolites). Samples for NMR-based metabolomics can be analyzed without extraction of metabolites provided that the active volume of the NMR probe is filled, typically with 500 µL of biofluid for 5 mm NMR tubes. However, the intrinsic low resolution of NMR greatly hinders accurate assignments and quantification of a large number of metabolites from biofluids[4,8,9] or intact tissues (i.e., ^{1}H-MAS NMR).[19–21] In this context, extraction of metabolites may reduce broad signals in the spectra arising from the high overlap of chemical shifts for metabolites and macromolecules (e.g., proteins) and generate narrower and better-resolved NMR resonances that allow reliable quantification of metabolites.[22] Samples for LC-MS-based metabolomics usually require

protein removal under nonphysiological conditions, such as high percentage of organic solvents and acidic pH, and injections of <20 µL of the extracted compounds.[23–27] The reproducibility of NMR measurements using these sample preparation procedures for LC-MS, however, can be severely affected by the pH and the ionic strength of the solution. Therefore, in the few studies that have aimed to measure the same sample (e.g., tissue extract, cell cultures, serum) by using LC-MS and NMR, the original sample was aliquoted and subjected to different extraction protocols.[28] Alternatively, the supernatant obtained after the extraction process is aliquoted, dried and later dissolved in specific solvents for NMR or LC-MS.[12] This common practice requires initial large sample volumes, which may preclude the analysis of important biofluids, such as vitreous humor or cerebrospinal fluid, and small tissues sections, such as those obtained from biopsies. Under these premises, there is a need in metabolomic studies for developing compatible extraction protocols, which enable direct measurement of metabolites both by NMR and LC-MS. In the present study, we use liver tissue to optimize metabolite extraction conditions for NMR analysis, followed by LC-ESI-MS by using the same sample extract with no need for solvent exchange or further pretreatment. Specifically, we applied 12 extraction methods involving different aqueous/organic solvents and three different temperatures to liver tissue. We investigated the number of metabolites, the extraction efficiency, and the reproducibility of the methods by using NMR and LC-MS.

EXPERIMENTAL SECTION

Reagents. LC-MS grade methanol (MeOH) and acetonitrile (ACN) and analytical grade chloroform ($CHCl_3$) were purchased from SDS (Peypin, France). Water was produced in an in-house Milli-Q purification system (Millipore, Molsheim, France). Ammonium acetate, formic acid, and ammonium fluoride were purchased from Sigma-Aldrich (Steinheim, Germany). Deuterated acetonitrile (CD_3CN) was purchased from Eurisotop, and deuterated water (D_2O) and 5-mm NMR tubes were purchased from Cortecnet (Viosins Le Bretonneux, France).

Sample Collection. The liver of a single rat was used to avoid biological variability in the comparative analysis. The rat was anaesthetized using sodium pentobarbital and exsanguinated via the abdominal aorta. Next, the liver was removed, immediately frozen in liquid nitrogen and kept at −80 °C until further process.

Metabolite Extraction Methods. Metabolites were extracted by using one of the following solvent combinations: MeOH/H2O (M/W) (1:1), ACN/H2O (A/W) (1:1), MeOH/CHCl3/H2O (M/C/W) (7:2:1), and ACN/CHCl3/H2O (A/C/W) (7:2:1) at three different temperatures: −20 °C, room temperature (RT) (25 °C), and 60 °C. The frozen liver was crushed in liquid N_2, and the pulverized tissue was collected to generate 12 aliquots of 100 mg. Two milliliters of each solvent system was added to the aliquots of liver tissue, and the mixture was further homogenized using a Teflon homogenizer (Black & Decker D-108-S1) for 1 min. Homogenization of the sample and extraction of metabolites using solvents at −20 °C and RT was performed on a water−ice bath to avoid overheating. Next, the sample was incubated 1 h at −20 °C, followed by a

15-min centrifugation at 4 °C and 13 000 rpm. The resultant supernatant was transferred to a separate vial, and the precipitate was extracted twice more following the same procedure, pooling together the supernatants to a final volume of ~6 mL.. Finally, the supernatant was evaporated to dryness under a stream of oxygen-free N_2 and redissolved in 500 µL of deuterated acetonitrile and deuterated water (2:8).

NMR Analysis. Redissolved samples in deuterated acetonitrile/water (2:8) were placed into 5 mm NMR tubes. ^{1}H NMR spectra were recorded at 300 K on an Avance III 600 spectrometer (Bruker, Germany) operating at a proton frequency of 600.20 MHz using a 5 mm CPTCI triple resonance (^{1}H, ^{13}C, ^{31}P) gradient cryoprobe. Given the high technical reproducibility of NMR, a single analysis was run for each sample. One-dimensional ^{1}H pulse experiments were carried out using the nuclear Overhauser effect spectroscopy (NOESY) presaturation sequence (RD−90°−t1−90°−tm−90° ACQ) to suppress the residual water peak, and the mixing time was set at 100 ms. Solvent presaturation with irradiaton power of 75 Hz was applied during recycling delay (RD = 5 s) and mixing time. The 90° pulse length was calibrated for each sample and varied from 6.57 to 6.99 ms. The spectral width was 12 kHz (20 ppm), and a total of 256 transients were collected into 64 k data points for each ^{1}H spectrum. The exponential line-broadening applied before Fourier transformation was of 0.3 Hz. The frequency domain spectra were phased and baseline-corrected using TopSpin software (version 2.1, Bruker).

HR-MAS Analysis. Frozen tissues samples (~15 mg) were placed in an insert for a 4 mm o.d. ZrO_2 rotor, limiting the rotor inner volume to 20

µL. The insert was filled with cooled D2O, sealed and subsequently inserted into the ZrO_2 rotor. HR-MAS spectra were recorded on a Bruker Avance III 500 spectrometer operating at a proton frequency of 500.13 MHz. The instrument was equipped with a 4 mm triple resonance (^{1}H, ^{13}C, ^{31}P) gradient HR-MAS probe. A Bruker Cooling Unit (BCU-Xtreme) was used to keep the sample temperature at 4 °C. Samples conveniently prepared with D_2O were spun at 4 kHz to keep the rotation sidebands out of the spectral region of interest. One-dimensional (1D) ^{1}H spectra were acquired, adding 128 scans in 23 min, with a 7 kHz (14 ppm) spectral width and 8.0 s of relaxation delay to ensure full relaxation of magnetization between scans. A ^{1}H HR-MAS spectra of low molecular weight metabolites was performed using the Carr Purcell Meiboom Gill sequence (cpmg spin.spin T2 relaxation filter) for identification purposes, with a total time filter of 171 ms that attenuated the macromolecule signals to a residual level. An additional 2-dimensional J-resolved spectrum (2D JRES) aided in the identification of multiplicity of the liver metabolite signals.

LC-MS Analysis. Portions (100 µL) of each redissolved sample in deuterated acetonitrile/water (2:8) were placed into HPLC vials after NMR analysis. LC-MS analyses were performed using an HPLC system (1200 series, Agilent Technologies) coupled to a 6230 ESI-QTOF (Agilent Technologies) operated in positive (ESI+) or negative (ESI-) electrospray ionization mode. Vials containing extracted metabolites using one of the 12 conditions described above were kept at 20 °C prior to LC-MS analysis. Each sample (i.e., vial) was run in triplicate. Metabolites were separated using a Waters XBridge column C18, 150 x 2.1 mm, 3.5 µm

(Waters, Ireland). When the instrument was operated in positive ionization mode, the solvent system was A1 = 0.1% formic acid in water and B1 = 0.1% formic acid in acetonitrile. When the instrument was operated in negative ionization mode, the solvent system was A2 = 1 mM ammonium fluoride in water and B2 = acetonitrile.[23] The linear gradient elution started at 100% A (time 0.3 min) and finished at 100% B (25.30 min). The injection volume was 10 uL. ESI conditions were gas temperature, 325 ºC; drying gas, 11 L min^{-1}; nebulizer, 40 psig; fragmentor, 120 V; and skimmer, 65 V. The instrument was set to acquire over the m/z range 80-1200 with an acquisition rate of 1.2 spectra/s. MS/MS was performed in targeted mode, and the instrument was set to acquire over the m/z range 50-1000, with a default iso width (the width half-maximum of the quadrupole mass bandpass used during MS/MS precursor isolation) of 4 m/z. The collision energy was fixed at 20 V.

Data Analysis. The acquired ^{1}H NMR spectra were phased, baseline-corrected, and referenced to the chemical shift of residual acetonitrile signal at 2.1 ppm. References of pure compounds from the metabolic profiling AMIX spectra database (Bruker), HMDB, and Chenomx databases were used for metabolite identification. In addition, we assigned metabolites by ^{1}H-^{1}H homonuclear correlation (COSY and TOCSY) and ^{1}H-^{13}C heteronuclear (HSQC) 2D NMR experiments and by correlation with pure compounds run inhouse, such as glucose, tyrosine, or creatinine, among others. After baseline correction, specific ^{1}H NMR regions identified in the spectra were integrated for each extraction method entering the study using the AMIX 3.8 software package. Data

(pre-)processing, data analysis, and statistical calculations were performed in Matlab (Matlab version 6.5.1, Release 13). LC-MS data from the liver extractions (ESI+ and ESI- modes) were processed by using the XCMS software[29] (version 2.9.2) to detect and align features. Samples were run as triplicates, and only features consistently present in all three runs were considered for quantification. XCMS analysis of these data provided a matrix containing the retention time, m/z value, and intensity of each feature for every extraction method discussed above.

RESULTS

Influence of Extraction Method on the Analysis of Metabolites by NMR.

The 12 different extraction protocols were assayed on the liver tissue obtained from the same rat (see Experimental Section). The selected protocols represent examples in which two fundamental conditions for metabolite solubility and stability are varied, such as solvent polarity and temperature. Each data set was scaled to unit variance and projected using an unsupervised principal component analysis (PCA) to show similarities in the data, and the resulting scores and loadings were visualized in a score-loading plot (biplot) (Figure 1).

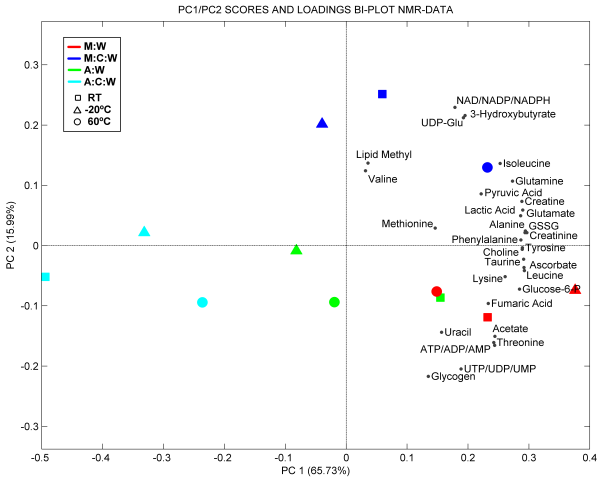


Figure 1. Biplot of NMR data. The biplot is a scatter plot that graphically displays loadings and scores from the principal component analysis (PCA) of the NMR data. It displays scores on the 12 extraction conditions and loading on the metabolites identified by NMR. The biplot shows that solvent combination (color code) has a greater effect in the extraction of metabolites than solvent temperature (shape code). M/W and M/C/W showed the best extraction yields for the 30 metabolites identified at three different temperatures: −20 °C, room temperature (RT) (25 °C), and 60 °C. M:W = MeOH/H2O; M:C:W = MeOH/CHCl3/H2O; A:W = ACN/H2O, and A:C:W = ACN/CHCl3/H2O.

Similar methods are near each other, and dissimilar methods are farther apart from each other. The biplot explains 81.7% of the total variance in the data and shows that the type of solvent (color code) used during the extraction of metabolites has a greater effect in the extraction efficiency than solvent temperature (shape code). The M/W protocol followed by M/C/W turned out to be the most efficient extraction protocols used in this study. The M/W protocol, in addition, appeared slightly less variable to temperature than M/C/W. In contrast, A/C/W proved to be the least efficient in extracting metabolites. To investigate

in detail the influence of the extraction protocol on the concentration of metabolites by NMR, we examined the concentration of 30 specific metabolites identified in liver as displayed in Figure 2. Metabolites characterized by different polarities and chemical functional groups were extracted with different efficiencies according to the method used.

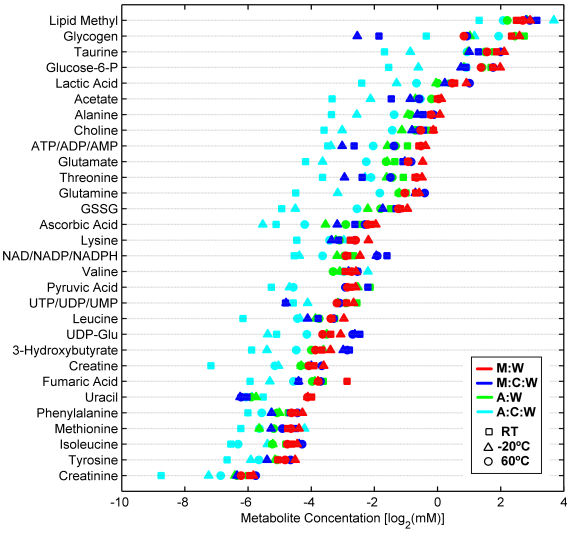


Figure 2. Quantification of 30 metabolites extracted from liver using 12 different methods and analyzed using ^{1}H NMR. The concentration (mM) scale of the X axis is log 2.

Among the endogenous metabolites characterized by using NMR-based metabolomics, we show in Figure 2 important primary metabolites of central metabolism, such as fumaric acid, lactic acid, 3-hydroxybutyrate, pyruvic acid, or taurine, as well as amino acids, cofactors, or glucose, among others. In addition to low-molecular-weight metabolites, another interesting capability of NMR for metabolomic studies is the detection of glycogen, a glucose-based polymer that serves as the secondary long-term energy storage in animals. Overall, when

chloroform is used as the extraction solvent in protocols M/C/W and A/C/W, temperature produces higher variability in the extraction efficiency of metabolites. This effect can be explained by the low boiling point of chloroform in the mixture. Importantly, from comparing HR-MAS and extraction methods for NMR from liver tissue, we conclude that resolution of liquid NMR and HR-MAS is similar for many aqueous metabolites. However, we detected a larger number of metabolites using extracted samples. Figure 3 shows enhanced detection of aqueous compounds by using the extraction method for NMR compared with HR-MAS of the same liver tissue.

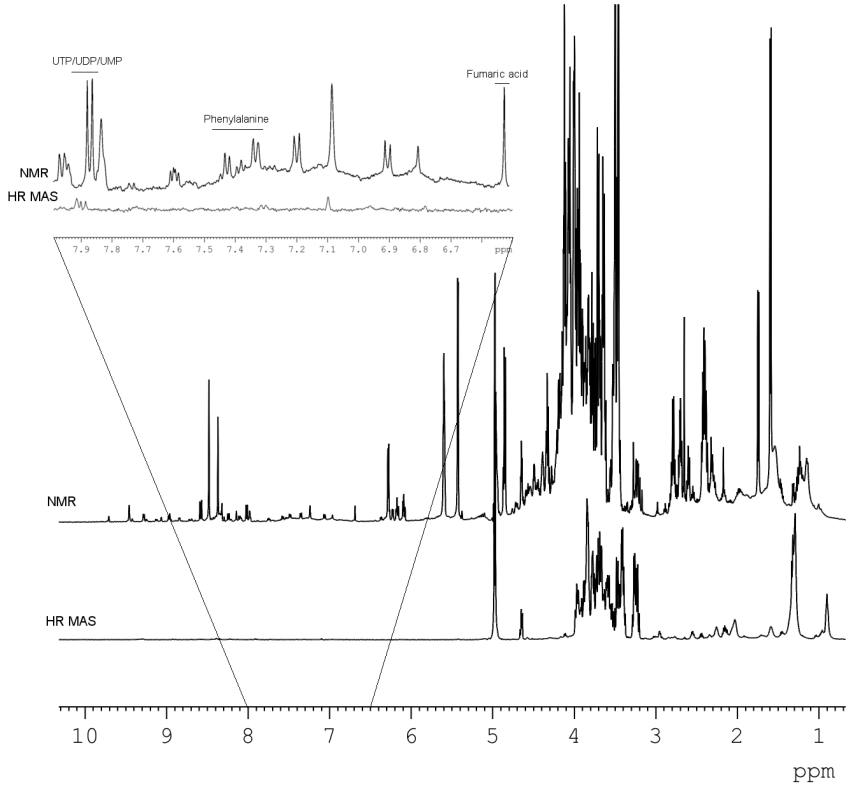


Figure 3. Top: NMR spectrum of liver extraction. Bottom: HRMAS spectrum from intact liver tissue.

LC-MS Analysis of Extracted Metabolites Reconstituted in Deuterated Solvents.

Direct LC-MS analysis of extracted metabolites reconstituted in deuterated solutions with no need to exchange the solvent would greatly facilitate the integration of NMR and MS technologies in metabolomics. However, the use of polar deuterated solvents can induce chemical exchange of acidic ^{1}H in metabolite structures for 2D in solution, which would seriously hamper the identification of such metabolites by using accurate mass information. To explore this issue, we performed LC-MS analysis of pure standard compounds and metabolites extracted from liver tissue reconstituted in either water (H_2O) or deuterated water (D_2O). Our results in Figure 4 indicate that neither the number nor the abundance of features from the liver extract differ when the samples are reconstituted in H_2O or D_2O (Figure 4A). The isotopic distribution of phenylalanine, tryptophan, and LysoPC (16:0) identified by MS/MS in the liver extract prove the absence of deuterium exchange during the LC-MS analysis (Figure 4B). The same phenomenon was observed using pure standard compounds containing amine groups (data not shown). Therefore, the requirement of deuterated solvents for NMR measurements does not prevent further analysis of the same extracted metabolites by LC-MS. Accordingly, all extractions previously analyzed using NMR were run in triplicate under the same LC-MS conditions (see Experimental Section). Thousands of features (unique m/z with a specific retention time) were detected, which translated into

300-700 [M - H]+ hits in METLIN database with an error of <5 ppm, depending on the extraction method used (Table 1).

Extraction Solvent	# Features > 5000 counts	Hits METLIN [MH^+] < 5ppm	% Hits METLN
A/W rt	2603	382	15%
A/W-20 °C	4647	695	15%
A/W60 °C	3517	546	16%
A/C/W rt	1365	301	22%
A/C/W-20 °C	3967	524	13%
A/C/W60 °C	2075	444	21%
M/W rt	1900	418	22%
M/W-20 °C	2371	493	21%
M/W60 °C	2303	433	19%
M/C/W rt	2048	532	26%
M/C/W-20 °C	2642	536	20%
M/C/W60 °C	1970	508	26%

Table 1. Number of features and hits in METLIN database for each extraction method

To illustrate the complementarity of LC-MS, we identified metabolites present at low concentrations that are not detectable by NMR. An unsupervised PCA analysis was conducted by scaling to unit variance the intensities of 22 selected metabolites identified by MS/MS data, and the resulting scores and loadings are visualized in Figure 5.

The biplot (Figure 5) covers 72.6% of the total variance and shows the dispersion of the extraction protocols and metabolite abundances according to their extraction efficiencies. The clustering of the LC-MS data confirms the NMR results displayed in Figure 1, showing that the main cause of variation in the extraction efficiency of metabolites is attributed to the combination of solvents (color code) rather than the temperature of these (shape code).

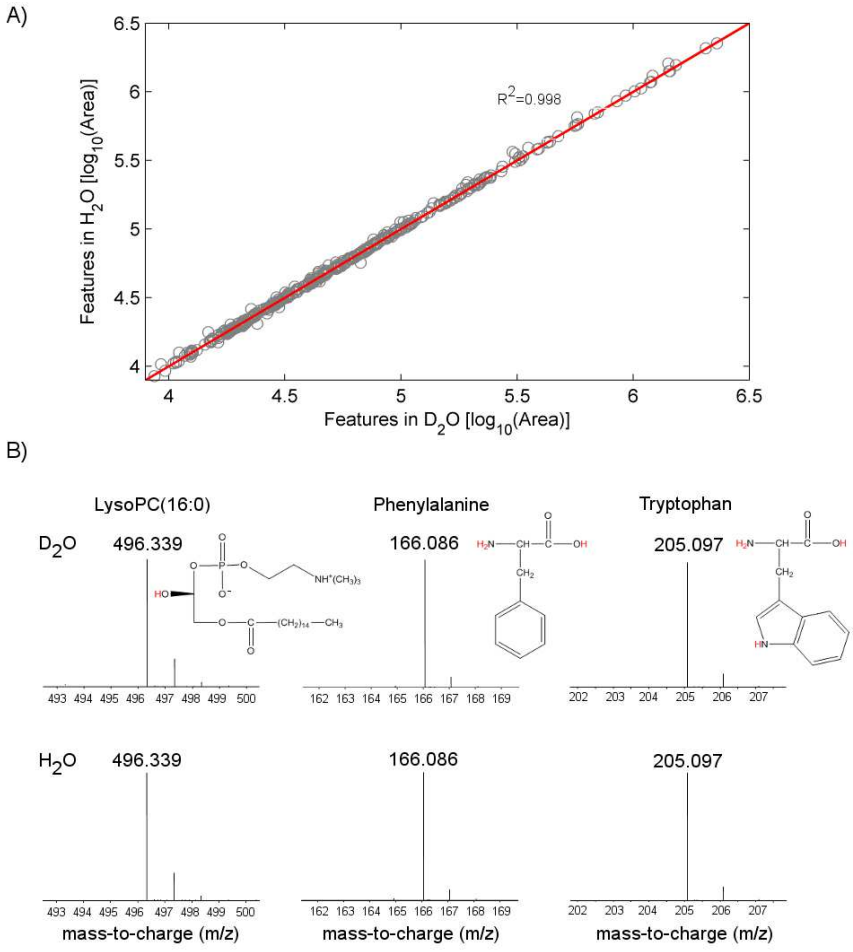


Figure 4. (A) Scatter plot representing the area of each feature from the XCMS matrix of LC-MS data of liver samples reconstituted in H_2O and D_2O. A correlation coefficient (R2) of 0.997 indicates a high linear regression, which demonstrates insignificant differences between the number and abundance of features detected in liver extracts reconstituted in H_2O and D_2O. (B) Mass spectra of phenylalanine, tryptophan, and LysoPC (16:0) reconstituted in D_2O (top) and H_2O (bottom). Labile hydrogens are marked in red. Mass spectra show that the isotopic distributions of the compounds are not altered by D_2O, indicating either poor exchange of acidic ^1H for 2D in solution or quick back-exchange of labile 2D in aqueous LC-MS buffers due to a total solvent accessibility of small molecule structures.

Again, methods containing methanol proved to be more efficient for metabolite extraction than those containing acetonitrile. For example, important cofactors such as coenzyme A (CoA); acetyl-CoA; propyonil-

CoA; malonil-CoA; and key oxidizing and reducing agents such as NAD+, NADH, and FAD were extracted more efficiently using M/C/W, whereas hypoxanthine, spermidine, and guanidine were more efficiently extracted with M/W. Overall, we interpret our results in decreasing order of efficiency in extracting metabolites and complementarity of NMR and MS analysis as follows: M/C/W > M/W > A/W > A/C/W.

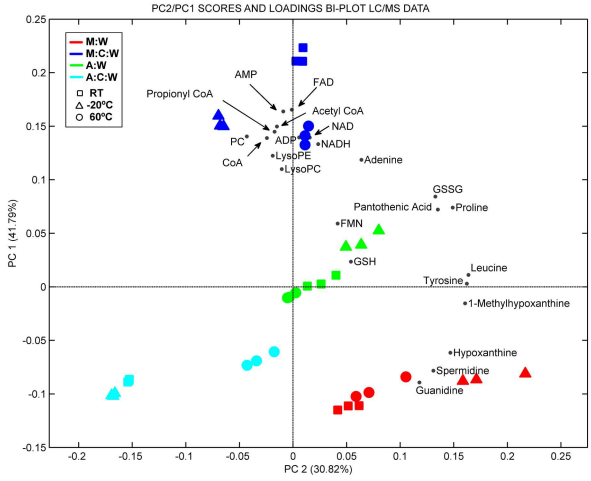


Figure 5. Biplot of LC-MS data. The biplot displays scores on the 12 extraction conditions and loading on the metabolites identified by MS/MS. Solvent combination (color code) has a greater effect in the extraction of metabolites than solvent temperature (shape code). M:W and M:C:W showed the best extraction yields for the 23 metabolites identified; however, important cofactors, such as coenzyme A (CoA); acetyl-CoA; propyonil-CoA; malonil-CoA; and key oxidizing and reducing agents, such as NAD+, NADH, and FAD, were extracted more efficiently using M:C:W. M:W = MeOH/H2O; M:C:W = MeOH/CHCl3/H2O; A:W = ACN/H2O, and A:C:W = ACN/ CHCl3/H2O at three different temperatures: −20 °C, room emperature, (RT) (25 °C), and 60 °C.

Reproducibility of the extraction method.

Finally, we assessed the reproducibility of the M/C/W protocol, the topranked method in our study. Technical variation arises mainly from sampling, metabolite extraction, and variability of the instrumental platform. Here, variation from sampling was diminished by using a single crushed liver tissue aliquoted into 12 fractions of 100 mg. To nearly eliminate instrumental variability, we used NMR measurements, since, from in-house experience, the coefficient of variation (% CV) from four consecutive NMR measurements of the same liver extract (i.e., repeatability) and four measurements of the same extract in four different days (i.e., reproducibility) plus manual metabolite fitting is <1% and <2%, respectively. In contrast, the repeatability of our LC-MS measurements is much higher, 15-20%, as previously reported by other groups.[30-32] Therefore, by using NMR, we ensured that the total variability could be principally attributed to the reproducibility of the extraction process. Metabolite extractions were then run in triplicate using the M/C/W protocol at -20 ºC, room temperature and 60 ºC. The percent CV of 30 metabolites shown in Figure 6 indicates that metabolites are extracted with high reproducibility showing a median CV of 13.8% at .20 ºC, 12.8% at RT, and 14.2% at 60 ºC. With the exception of glycogen, multiphosphorylated nucleotides and ascorbic acid (discussed below), the temperature of the solvent has a minor effect on the variability of the extraction. This variation is slightly higher than the average variability of 7% CV in HR-MAS that we elucidated by measuring the same liver tissue five times.

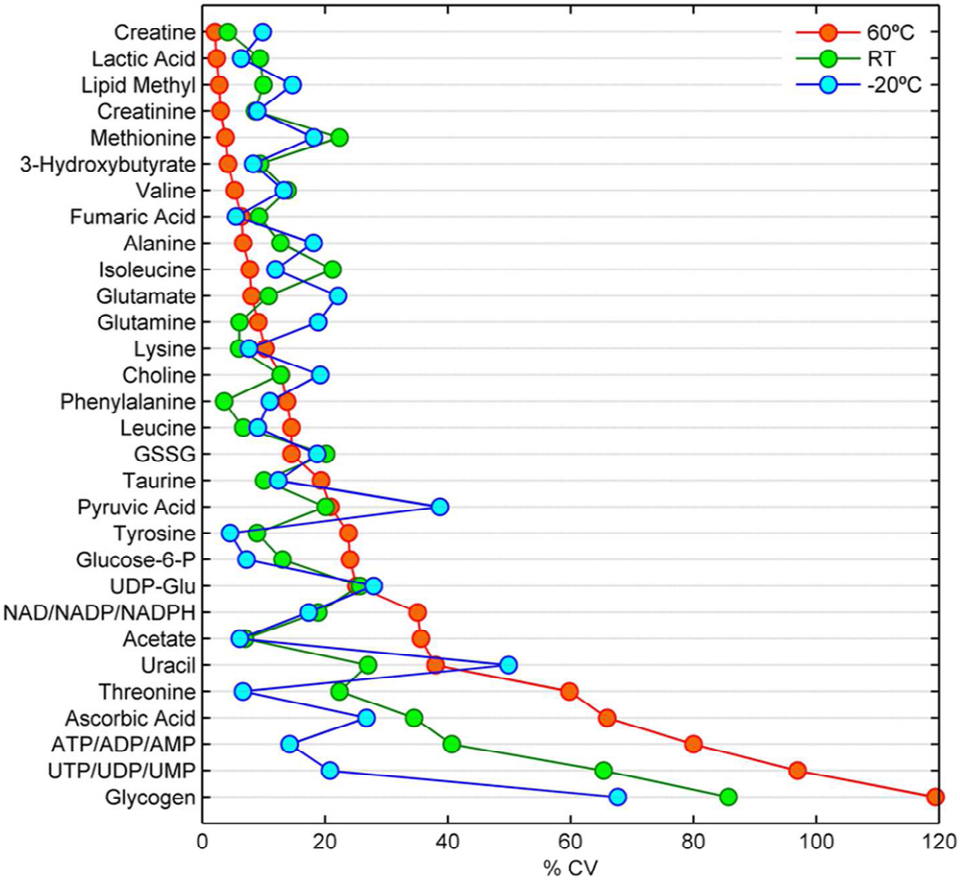


Figure 6. Coefficient of variation for metabolites extracted using the M/C/W protocol at −20 °C, room temperature, and 60 °C. Metabolite extractions were then run in triplicate at each temperature, and data points represent the percent CV of metabolite concentration analyzed using 1H NMR. Mean CV at −20 °C is 13.8%, 12.8% at RT, and 14.2% at 60 °C.

DISCUSSION

The distinctive physicochemical properties measured by NMR and LC-MS make these two technologies highly complementary for metabolomic studies. Thus, the current study aimed to explore optimal and complementary extraction conditions for NMR and LC-MS analysis since, to the best of our knowledge, this has been an unexplored issue in metabolomics. Importantly, our experimental design was conditioned by the self-imposed requirement that NMR could analyze extracted metabolites first and the same mixture analyzed by LC-MS afterward with no need for solvent exchange. The major limitation of this approach could arise from the use of deuterated solvents for NMR measurements, since incorporation of 2D into labile ^{1}H positions of metabolites could impede the correct identification of these compounds by MS. However, the analysis by LC-MS of the liver extract dissolved in D_2O proved to be satisfactory, and we did not observe deuterium exchange in either the LC-MS features or the standard metabolites interrogated. We therefore interpret these results such that the interaction of the chromatographic mobile phase with the injected sample (dissolved in D_2O) results in a complete and quick reversion (i.e., back-exchange) of labile 2D positions back to ^{1}H in the presence of aqueous buffers since, unlike proteins, metabolite structures are totally solvent-exposed. Therefore, to assess metabolite extraction efficiencies, we used four solvent compositions and three temperatures. The solvents used were combinations of methanol, acetonitrile, water, and chloroform. On the basis of previous study by Zhang et al.,[33] the percentage of chloroform in the extraction solvent was set at 20% to avoid phase separation. We avoided

metabolite extraction conditions typically used for LC-MS based metabolomics characterized by extreme pH or high salt content, given the susceptibility of NMR measurements to the ionic strength and pH of a solution that may result in signal shifting, peak broadening, lower reproducibility and difficulties in instrument acquisition. Solvent temperatures were selected on the basis of the effect that temperature has in the solubility of metabolites. Finally, we used liver as a model tissue because this organ plays a major role in metabolism, regulating a wide variety of biochemical reactions, including the synthesis of amino acids, glucose, and lipids and the breakdown of glycogen and toxic substances. Consequently, such chemical and structural diversity of metabolites makes this tissue very attractive for the aim of the study. The liver is also prone to common diseases, such as hepatitis, alcohol damage, fatty liver, cirrhosis, cancer, or drug damage, and given its role in glucose homeostasis, the liver may cause or contribute to diabetes mellitus. The results of our study showed that the choice of solvents largely determined extraction efficiencies, and extraction methods based on methanol/chloroform/water (7/2/1) and methanol/water (1/1) best showed the complementarity of NMR and LC-MS in metabolomics, providing high extraction yields and minimal repetition of metabolites between platforms. In contrast, the temperature of the solvent mixture had a very minor effect, with few exceptions discussed below. In this context, it is important to note the consistency of our results between the two analytical platforms used to evaluate the data. Our results are in agreement with Geier et al., [34] who observed that the choice of solvent had more influence than any other consideration tested, such as the

tissue disruption method. In our study, metabolites difficult to detect using untargeted reverse-phase LC-MS, such as glucose, cholesterol, ascorbic acid, pyruvic acid, or lactic acid, were easily analyzed using NMR. One remarkable example is glycogen, a polymer whose function as a glucose reservoir makes its reliable quantification particularly important in diseases such as diabetes. Because of its polymeric nature, glycogen cannot be ionized by MS, whereas its quantification by NMR is relatively simple. It is worth mentioning, however, the high extraction variability of glycogen, particularly at high temperatures, which we attribute to its extremely labile nature and its heterogenic distribution across the liver.[35] We also noticed that elevated temperature has a detrimental effect on the detection of multiphosphorylated species, such as ATP and UTP, and chemically active and unstable compounds, such as ascorbic acid, which we attribute to their degradation and oxidation, respectively. Nevertheless, the median CV of 13% for 30 extracted metabolites quantified by NMR is comparable to HR-MAS measurements of intact liver tissue. Given the small dimensions of MAS rotors, technical variability in HR-MAS is also predominantly influenced by sample preparation. Blaise et al.[36] estimated a maximum variance of 21% for rotor filling by comparing MAS rotors lightly vs heavily loaded. In addition, we estimated that the technical variability as a result of instrument shimming for HR-MAS was 7% CV, which is in accordance with the variation of 5-10% reported by Blaise et al.[36] This variability could be explained by the manual manipulation of the rotor and by the possible degradation or metabolic transformation of metabolites. Altogether, we recognize the utility of HR-MAS for targeted applications

in metabolomics; however, when attempting large sample batches in real metabolomic studies for biomarker discovery, we consider that sample handling and instrument shimming [37] for HR-MAS is labor-intensive and extremely time-consuming compared with NMR analysis from liquid extracts that enable complete automation. Our results also highlight another interesting feature of methanol-containing methods in that protein removal is more effective than those containing acetonitrile, as revealed by the reduced number and intensity of multiply charge species attributed to large peptides and proteins in the mass spectra. This is interesting, since the presence of proteins in the sample can suppress the signal of metabolites eluting at similar retention times and hamper the chromatographic performance. This result is in concordance with Want et al.[38] who reported that methanol-based extraction methodologies efficiently precipitate proteins from serum. In contrast, Rabinowitz and colleagues seem to have a preference for acidic acetonitrile as extraction solvent for metabolomics.[39] In our study, however, the use of acidic conditions was avoided, given the susceptibility of NMR measurements, which may result in signal shifting, peak-broadening, and lower reproducibility. In summary, since our results prove that deuterated solvents used for NMR have essentially no effect on the MS data, the exploration of compatible extraction procedures for both NMR and MS-based analyses opens up new opportunities to expand coverage of the metabolome by implementing multiplatform approaches based on NMR and MS. Yet, we acknowledge that the difficulty of establishing the “best extraction procedure” for global metabolomics will remain in the field.

BIBLIOGRAPHY

(1) Yanes, O.; Clark, J.; Wong, D. M.; Patti, G. J.; Sanchez-Ruiz, A.; Benton, H. P.; Trauger, S. A.; Desponts, C.; Ding, S.; Siuzdak, G. Nat. Chem. Biol. 2010, 6, 411-417.

(2) Patti, G. J.; Yanes, O.; Shriver, L. P.; Courade, J. P.; Tautenhahn, R.; Manchester, M.; Siuzdak, G. Nat. Chem. Biol. 2012, DOI/ 10.1038/nchembio.767.

(3) Koves, T. R.; Ussher, J. R.; Noland, R. C.; Slentz, D.; Mosedale, M.; Ilkayeva, O.; Bain, J.; Stevens, R.; Dyck, J. R. B.; Newgard, C. B.; Lopaschuk, G. D.; Muoio, D. M. Cell Metab. 2008, 7, 45-56.

(4) Vinaixa, M.; Rodriguez, M. A.; Samino, S.; Diaz, M.; Beltran, A.; Mallol, R.; Blade, C.; Ibanez, L.; Correig, X.; Yanes, O. PLoS One 2011, 6, e29052.

(5) Wishart, D. S. Bioanalysis 2011, 3, 1769-1782.

(6) Want, E. J.; Wilson, I. D.; Gika, H.; Theodoridis, G.; Plumb, R. S.; Shockcor, J.; Holmes, E.; Nicholson, J. K. Nat. Protoc. 2010, 5, 1005-1018.

(7) De Meyer, T.; Sinnaeve, D.; Van Gasse, B.; Tsiporkova, E.;Rietzschel, E. R.; De Buyzere, M. L.; Gillebert, T. C.; Bekaert, S.;Martins, J. C.; Van Criekinge, W. Anal. Chem. 2008, 80, 3783-3790.

(8) Verwaest, K. A.; Vu, T. N.; Laukens, K.; Clemens, L. E.; Nguyen, H. P.; Van Gasse, B.; Martins, J. C.; Van Der Linden, A.; Dommisse, R. Biochim. Biophys. Acta 2011, 1812, 1371-1379.

(9) Amathieu, R.; Nahon, P.; Triba, M.; Bouchemal, N.; Trinchet, J.C.; Beaugrand, M.; Dhonneur, G.; Le Moyec, L. J. Proteome Res. 2011, 10, 3239-3245.

(10) Ciborowski, M.; Martin-Ventura, J. L.; Meilhac, O.; Michel, J. B.; Ruperez, F. J.; Tunon, J.; Egido, J.; Barbas, C. J. Proteome Res. 2011, 10, 1374-1382.

(11) Patti, G. J. J. Sep. Sci. 2011, 34, 3460-3469.

(12) Shaham, O.; Slate, N. G.; Goldberger, O.; Xu, Q. W.; Ramanathan, A.; Souza, A. L.; Clish, C. B.; Sims, K. B.; Mootha, V. K. P. Natl. Acad. Sci. USA 2010, 107, 1571-1575.

(13) Wikoff, W. R.; Anfora, A. T.; Liu, J.; Schultz, P. G.; Lesley, S. A.; Peters, E. C.; Siuzdak, G. P. Natl. Acad. Sci. USA 2009, 106, 3698-3703.

(14) Nevedomskaya, E.; Mayboroda, O. A.; Deelder, A. M. Mol. Biosyst. 2011, 7, 3214-3222.

(15) Goodpaster, A. M.; Ramadas, E. H.; Kennedy, M. A. Anal. Chem. 2011, 83, 896-902.

(16) Wishart, D. S.; Lewis, M. J.; Morrissey, J. A.; Flegel, M. D.; Jeroncic, K.; Xiong, Y. P.; Cheng, D.; Eisner, R.; Gautam, B.; Tzur, D.; Sawhney, S.; Bamforth, F.; Greiner, R.; Li, L. J. Chromatogr. B 2008,

871, 164-173.

(17) Moco, S.; Forshed, J.; De Vos, R. C. H.; Bino, R. J.; Vervoort, J. Metabolomics 2008, 4, 202-215.

(18) Lanza, I. R.; Zhang, S. C.; Ward, L. E.; Karakelides, H.; Raftery, D.; Nair, K. S. PLoS One 2010, 5, e10538. DOI/ 10.1371/journal.pone.0010538.

(19) Beckonert, O.; Coen, M.; Keun, H. C.; Wang, Y.; Ebbels, T. M.; Holmes, E.; Lindon, J. C.; Nicholson, J. K. Nat. Protoc. 2010, 5, 1019-1032.

(20) Li, M.; Song, Y.; Cho, N.; Chang, J. M.; Koo, H. R.; Yi, A.; Kim, H.; Park, S.; Moon, W. K. PLoS One 2011, 6, e25563.

(21) Monleon, D.; Morales, J. M.; Gonzalez-Darder, J.; Talamantes, F.; Cortes, O.; Gil-Benso, R.; Loezin-Gines, C.; Cerda-Nicolas, M.;Celda, B. J. Proteome Res. 2008, 7, 2882-2888.

(22) Vinaixa, M.; Rodriguez, M. A.; Rull, A.; Beltran, R.; Blade, C.;Brezmes, J.; Canellas, N.; Joven, J.; Correig, X. J. Proteome Res. 2010, 9, 2527-2538.

(23) Yanes, O.; Tautenhahn, R.; Patti, G. J.; Siuzdak, G. Anal. Chem. 2011, 83, 2152-2161.

(24) Gonzalez, B.; Francois, J.; Renaud, M. Yeast 1997, 13, 1347-1355.

(25) Canelas, A. B.; ten Pierick, A.; Ras, C.; Seifar, R. M.; van Dam, J. C.; van Gulik, W. M.; Heijnen, J. J. Anal. Chem. 2009, 81, 7379-7389.

(26) Villas-Boas, S. G.; Hojer-Pedersen, J.; Akesson, M.; Smedsgaard, J.; Nielsen, J. Yeast 2005, 22, 1155-1169.

(27) Maharjan, R. P.; Ferenci, T. Anal. Biochem. 2003, 313, 145-154.

(28) Ruiz-Aracama, A.; Peijnenburg, A.; Kleinjans, J.; Jennen, D.; van Delft, J.; Hellfrisch, C.; Lommen, A. BMC Genom. 2011, 12, 251-270.

(29) Smith, C. A.; Want, E. J.; O'Maille, G.; Abagyan, R.; Siuzdak, G. Anal. Chem. 2006, 78, 779-787.

(30) Masson, P.; Alves, A. C.; Ebbels, T. M.; Nicholson, J. K.; Want, E. J. Anal. Chem. 2010, 82, 7779-7786.

(31) Crews, B.; Wikoff, W. R.; Patti, G. J.; Woo, H. K.; Kalisiak, E.; Heideker, J.; Siuzdak, G. Anal. Chem. 2009, 81, 8538-8544.

(32) Benton, P. H.; Want, E. J.; Clayton, T. A.; Keun, H. C.; Amberg, A.; Plumb, R. S.; Goldfain-Blanc, F.; Walther, B.; Reily, M. D.; Lindon, J. C.; Holmes, E.; Nicholson, J. K.; Ebbels, T. M. Anal. Chem. 2012, 84, 2424-2432.

(33) Zhang, Y.; Jiye, A.; Wang, G. J.; Huang, Q.; Yan, B.; Zha, W. B.; Gu, S. H.; Liu, L. S.; Ren, H. C.; Ren, M. T.; Sheng, L. S. J. Chromatogr., B 2009, 877, 1751-1757.

(34) Geier, F. M.; Want, E. J.; Leroi, A. M.; Bundy, J. G. Anal. Chem. 2011, 83, 3730-3736.

(35) Kudryavtseva, M. V.; Bezborodkina, N. N.; Okovity, S. V.; Kudryavtsev, B. N. Eur. J. Gastroenterol. Hepatol. 2001, 13, 693-697.

(36) Blaise, B. J.; Giacomotto, J.; Elena, B.; Dumas, M. E.; Toullhoat, P.; Segalat, L.; Emsley, L. Proc. Natl. Acad. Sci. USA 2007, 104, 19808-19812.

(37) Piotto, M.; Elbayed, K.; Wieruszeski, J. M.; Lippens, G. J. Magn. Reson. 2005, 173, 84-89.

(38) Want, E. J.; O'Maille, G.; Smith, C. A.; Brandon, T. R.; Uritboonthai, W.; Qin, C.; Trauger, S. A.; Siuzdak, G. Anal. Chem. 2006, 78, 743-752.

(39) Rabinowitz, J. D.; Kimball, E. Anal. Chem. 2007, 79, 6167−6173.

2.1.2 A Guideline to Univariate Statistical Analysis for LC-MS-Based Untargeted Metabolomics-Derived Data

SUMMARY

Several metabolomic software programs provide methods for peak picking, retention time alignment and quantification of metabolite features in LC-MS-based metabolomics. Statistical analysis, however, is needed in order to discover those features significantly altered between samples. By comparing the retention time and MS/MS data of a model compound to that from the altered feature of interest in the research sample, metabolites can be then unequivocally identified. This paper reports on a comprehensive overview of a workflow for statistical analysis to rank relevant metabolite features that will be selected for further MS/MS experiments. We focus on univariate data analysis applied in parallel on all detected features. Characteristics and challenges of this analysis are discussed and illustrated using four different real LC-MS untargeted metabolomic datasets. We demonstrate the influence of considering or violating mathematical assumptions on which univariate statistical test rely, using high-dimensional LC-MS datasets. Issues in data analysis such as determination of sample size, analytical variation, assumption of normalityand homocedasticity, or correction for multiple testing are discussed and illustrated in the context of our four untargeted LC-MS working examples.

1. Introduction

The comprehensive detection and quantification of metabolites in biological systems, coined as 'metabolomics', offers a new approach to interrogate mechanistic biochemistry related to natural processes such as health and disease. Recent developments in MS and NMR have been crucial to facilitate the global analysis of metabolites. The examination of metabolites, however, commonly follows two strategies: (i) targeted metabolomics, driven by a specific biochemical question or hypothesis in which a set of metabolites related to one or more pathways are defined, or (ii) untargeted metabolomics, driven by an unbiased approach (*i.e.*, nonhypothesis) in which as many metabolites as possible are measured and compared between samples [1]. The latter is comprehensive in scope and outputs complex data sets, particularly by using LC-MS-based methods. Thousands of so called metabolite features (*i.e.*, peaks corresponding to individual ions with a unique mass-to-charge ratio and a unique retention time or mzRT features from now on, can be routinely detected in biological samples. In addition, each mzRT feature in the dataset is associated with an intensity value (or area under the peak), which indicates its relative abundance in the sample. Overall, this complexity imposes the implementation of metabolomic softwares such as XCMS [2], MZmine [3] or Metalign [4] that can provide automatic methods for peak picking, retention time alignment to correct experimental drifts in instrumentation, and relative quantification. As a result, the identification of mzRT features that are differentially altered between sample groups has become a relatively automated process. However, the identification and quantization of a "metabolite feature"

does not necessary translate into a metabolite entity. LC-MS metabolomic data presents high redundancy because of the recurrent detection of adducts (Na+, K+, NH3, etc), isotopes, or doubly charged ions that greatly inflate the number of detected peaks. Several recently launched open-source algorithms such as CAMERA [5] or AStream [6], and commercially available software such as Mass Hunter (Agilent Technologies) or Sieve (Thermo Scientific), are capable of filtering redundancy by annotating isotopes and adduct peaks, and the resulting accurate compound mass (*i.e.*, molecular ion) can be searched in metabolite databases such as METLIN, HMDB or KEGG. Database matching represents only a putative metabolite assignment that must be confirmed by comparing the retention time and/or MS/MS data of a model pure compound to that from the feature of interest in the research sample. These additional analyses are time consuming and represent the rate-limiting step of the untargeted metabolomic workflow. Consequently, it is essential to prioritize the list of mzRT features from the raw data that will be subsequently identified by RT and/or MS/MS comparison. Relevant mzRT features for MS/MS identification are typically selected based on statistics criteria, either by multivariate data analysis or multiple independent univariate tests.

The intrinsic nature of biological processes and LC-MS-derived datasets is undoubtedly multivariatesince it involves observation and analysis of more than one variable at a time. Consequently, the majority of metabolomics studies make use of multivariate models to report their main findings. Despite the conferred utility, powerfulness and versatility of multivariate models, their performance might be fraught by the high-

dimensionality of such datasets due to the so-called 'curse of dimensionality' problem. Curse of dimensionality arises when datasets contain too much sparse data in terms of the number of input variables. This causes, in a given sample size, a maximum number of variables above which the performance of our multivariate model will degrade rather than improve. Hence, attempting to make the model conform too closely to this data (*i.e.*, considering too many variables in our multivariate model) can introduce substantial errors and reduce its predictive power (*i.e.*, overfitting). Therefore, using multivariate models require intensive validation work. Overall, multivariate data analysis is far from the scope of this paper and excellent reviews on multivariate tools for metabolomics can be found elsewhere [7,8]. On the other hand, data analysis can also be approached from a univariate perspective using traditional statistical methods that consider only one variable at a time [9]. The implementation of multivariate and univariate data analysis is not mutually exclusive and in fact, we strongly recommend their combined use to maximize the extraction of relevant information from metabolomic datasets [10,11]. Univariate methods are sometimes used in combination with multivariate models as a filter to retain those potentially "information-rich" mzRT features [12]. Then, the number of mzRT features considered in the multivariate model is significantly reduced down to those showing statistical significance in previous univariate tests (e.g., p-value < 0.05). On the other hand, there are multiple reported metabolomics works using univariate tests applied in parallel across all the detected mzRT features to report their main findings. It should be note that this approach overlooks correlations

within mzRT features and therefore information about correlated trends is not retained. In addition, applying multiple univariate tests in parallel to multivariate datasets involves the acceptance of mathematical pre-requisites and certain consequences such as the particular distributions of variables (e.g., normality) and increased risk of false positive results, respectively. Many researchers often ignore these issues when analyzing untargeted metabolomic datasets using univariate methods, which eventually can compromise their results.

This paper aims to investigate the impact of univariate statistical issues on LC-MS-based metabolomic experiments, particularly in small, focused studies (e.g., small clinical trials or animal studies). To this end, here we explore the nature of four real and independent datasets, evaluate the challenges and limitations of executing multiple univariate tests and illustrate available shortcuts. Note that we do not aim at writing a conventional statistical paper. Instead, our goal is to offer a practical guide with resources to overcome the challenges of multiple univariate analysis for untargeted metabolomic data. All methods described in this paper are based on scripts programmed either in MATLAB™ (Mathworks, Natick, MA) or R [13].

2. Properties of LC-MS Untargeted Datasets: High-Dimensional and Multicolinear

Basic information about the four real untargeted metabolomics LC-MS-based working examples is summarized in Table 1. These examples do not resemble ideal datasets described in basic statistical textbooks, and illustrates the challenges of real-life metabolomic experiments.

Working examples constitute retinas, serum and neuronal cell cultures under different experimental conditions (e.g., KO *vs*. WT; normoxia *vs*. hypoxia; treated *vs*. untreated) analyzed by LC-qTOF MS. Data were processed using the XCMS software to detect and align features, and thousands of features were generated from these biological samples. Each mzRT feature corresponds to a detected ion with a unique mass-tocharge ratio, retention time and raw intensity (or area). For example, each sample in example #3 exists in a space defined by 9877 variables or mzRT features. The four examples illustrate the high-dimensionality of untargeted LC-MS datasets in which the number of features or variables largely exceeds the number of samples. The rather limited number of individuals or samples per group is a common trait of metabolomic studies devoted to understand cellular metabolism [14-16]. When working with animal models of disease, for instance, this limitation is typically imposed by ethical and economical restrictions.

	Biofluid/Tissue	Sample groups	# samples /group	# XCMS variables	System	Reference
Example #1	Retina	KO	11	4581	LC/ESI-QTOF	[17]
		WT	11			
Example #2	Retina	Hypoxia	12	8146	LC/ESI-QTOF	[16]
		Normoxia	13			
Example #3	Serum	Untreated	12	9877	LC/ESI-TOF	[18]
		Treated	12			
Example #4	Neuronal cell cultures	KO	15	8221	LC/ESI-QTOF	unpublished data
		WT	11			

Table 1. Summary of working examples obtained from LC-MS untargeted metabolomic experiments. Further experimental details and methods can be obtained from references.(KO=Knock-Out; WT=Wild-Type)

Additionally, a second attribute of untargeted LC-MS metabolomic datasets is that they enclose multiple correlations among mzRT features

(*i.e.*, multicollinearity) [19]. Each metabolite produces more than one mzRT feature that result from isotopic distributions, potential adducts, and in-source fragmentation. Moreover, the evident biochemical interrelation among metabolites may also contribute to the multicollinearity. Namely, many metabolites participate in inter-connected enzymatic reactions and pathways (e.g., substrate and product; cofactors) and regulate enzymatic reactions (e.g., feed-back inhibition). Altogether, untargeted LC-MS metabolomics datasets are highly-dimensional and multicorrelated.

3. Sample Size Calculation in LC-MS Untargeted Metabolomics Studies

The number of subjects per group (*i.e.*, sample size) is an important aspect to be determined during the experimental design of the study. A low sample size may lead to a lack of precision, which may fail to provide reliable clues about the biological question under investigation. In contrast, an unnecessarily high sample size may lead to a waste of resources for minimal information gain. Thus it is not surprising that funding agencies require power/sample size calculations in their grant proposals. However, choosing the appropriate sample size for high-throughput approaches involving multivariate data is complicated. According to Hendriks *et al.* [8], there is currently nothing available for *a priori*sample size estimation of highly collinear multivariate data.

Traditional univariate sample size determination is based in the concept of power analysis. Power, or the sensitivity of the test, is defined as 1-β, being β the chance of a false negative or Type II error in

hypothesis testing. A Type II error is produced when a variable is claimed to not be significant when in fact it is. Therefore, power can be defined as the probability of a statistical test to allow detection of significant differences above a certain confidence. Classical power analysis to determine minimum sample size for a given variable (*i.e.*, metabolite) requires the estimation of population means and standard deviations and effect sizes. However, for high-dimensional data such estimates need to be redefined. Average power is used instead of power, significance level needs to take multiple testing into account and both effect sizes and variances take multiple values. Ferreira *et al*. [20,21] extended the concept of power analysis to high-dimensional data using univariate approaches in combination with multiple testing corrections. They used the entire set of test statistics from microarray pilot data to estimate the effect size distribution, power and minimal sample size. This method have been recently generalized and adapted by van Iterson *et al.* [22] as a part of the BioConductor package SSPA. Recall that using this method, data is treated as a set of multiple univariate responses and correlations between variables are ignored. On the other hand, this method was designed to guide experimental design decisions based on previously acquired pilot data. However, how realistic is to perform a pilot untargeted metabolomics study to determine minimum sample size? In practice, ethical and economical restrictions mainly determine the number of samples (*i.e.*, animals) for each group.

Although we recognize the limitations and controversy of post-hoc power analysis, for illustrative purposes we used SSPA to estimate effect sizes and perform power calculations of our untargeted metabolomics

data. Figure 1A show a comparison of example #2 and example #4 estimated power values considering up to 30 samples per group. Considering example #2, a 70% power to detect hypoxia-induced metabolic differences was obtained with 10 retinas per groups. This power was associated with a markedly bimodal density of effects sizes (Figure 1B) indicating significant hypoxiainduced metabolic variation. The density of effects sizes describes the effects observed in the data. Usually, a bimodal density is observed when the studied effect induces significant differences. In contrast, even considering up to 30 samples per group we end-up with low power to detect Koinduced differences in example#4 (Figure 1C). This indicates that KO-induced effects are scarcely reflected in our metabolomics data as represented by its unimodal densities of effects sizes Accordingly, we would estimate a minimum of ten samples per group (n = 10) as the easiest way to boost the statistical power of univariate statistical tests when true metabolic differences exist between two groups (e.g., example #2 comparing normoxia *vs*. hypoxia). This post-hoc calculation of the statistical power and sample size could be taken as a rough estimation for follow-up validation studies using QqQ instrumentation.

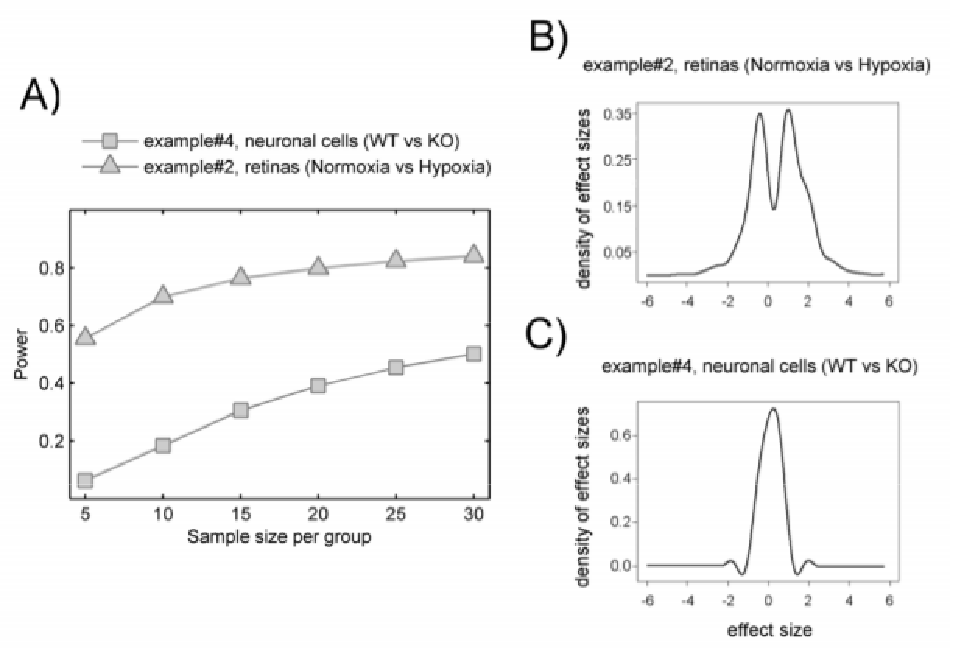


Figure 1. (A) Power curves for example #2 (Δ) and example #4 (☐) with sample size on the x-axis and estimated power using 5% FDR on the y-axis. Estimated densities of effect sizes for example #4 (B) and example #2 (C) with the standardized effect size on x-axis and estimated densities on the y-axis. Bimodal densities as in example #2 reflect more pronounced effects.

4. Handling Analytical Variation

The first issue that must be resolved before considering any univariate statistical test on LC-MS untargeted metabolomic data is analytical variation. Most common sources of analytical variation in LC-MS experiments are due to sample preparation, instrumental drifts caused by chromatographic columns and MS detectors, and errors caused in data processing [23].

The ideal method to examine analytical variation is to analyze QC samples, which will provide robust quality assurance of each detected mzRT feature [24]. To this end, QC samples should be prepared by

pooling aliquots of each individual sample and analyze them periodically throughout the sample work list. The performance of the analytical platform for each detected mzRT feature in real samples can be assessed by calculating the relative standard deviation of these features on pooled samples (CVQC) according to formula Equation (1), where S and X are respectively the standard deviation and the mean of each individual feature detected across the QC samples:

$$CV_{QC}(\%) = \frac{S}{\bar{X}} \times 100 \quad (1)$$

Likewise, the relative standard deviation of these features on study samples (CVT) can be defined according to formula Equation (2), where S and X are the standard deviation and mean respectively calculated for each mzRT feature across all study samples in the dataset.

$$CV_{T}(\%) = \frac{S}{\bar{X}} \times 100 \quad (2)$$

The variation of QC samples around their mean (CVQC) is expected to be low since they are replicates of the same pooled samples. Therefore Dunn *et al*. [24] have established a quality criteria by which any peak that presents a CVQC > 20% is removed from the dataset and thus ignored in subsequent univariate data analyses. Red and green spots in Figure 2 illustrate the CVT and CVQC frequencies distributions respectively for example #3 in which QC samples were measured. As expected, the highest percentage of mzRT features detected across QC samples present the lowest variation in terms of CVQC (green line). Conversely, the highest percentage of the mzRT features detected across the study samples holds the highest variation in terms of CVT (red line).

Notice that the intersection of red and green lines is produced around the threshold proposed by Dunn *et al.* [24]. Additionally, other studies performed on cerebrospinal fluid, serum or liver QC extracts also reported around 20% of CV on experimental replicates [25,26].

On the other hand, it is common that the nature of some biological samples and their limited availability complicates the analysis of QC samples. This was the case of mouse retinas in examples #1 and #2. Under these circumstances, there are not consensus standard criteria on how to handle analytical variation. We partially circumvent this issue using the following argument/ Provided that the total variation of a metabolite feature (CVT) can be expressed as a sum of biological variation (CVB) and analytical variation (CVA) according to Equation (3), computed CVT should be at minimum larger than 20% (the most accepted analytical variation threshold) for a metabolite feature to comprise biological variation.

$$CV_T^2 = CV_A^2 + CV_B^2 \quad (3)$$

Therefore, when QC samples are not available we propose as rule of thumb to discard those features showing CVT < 20% since biological variation is bellow analytical variation threshold. Figure 2 shows the frequency distribution of CVT for working examples #1,2 and #4 where QC samples were not available. According to our criteria, those mzRT features to the left of the threshold will hold more analytical than biological variation and should be conveniently removed from further statistical analysis. This surely results in a too broad criterion since it assumes that the analytical variation of all metabolites is similar, which is of course not accurate given that instrumental drifts do not affect all

metabolites evenly. It should be beard in mind, however, that tightly regulated metabolites presenting low variation such as glucose will likely be missed according to a 20% CVT cut-off criterion. Of mention, example #2 and example #4 show the higher and lower percentage of mzRT features with more than 50% CVT respectively. Therefore, there is more intrinsic variation in example #2 than in example#4. Whether such variation relates to the phenomena under study remain to be ascertained using hypothesis testing.

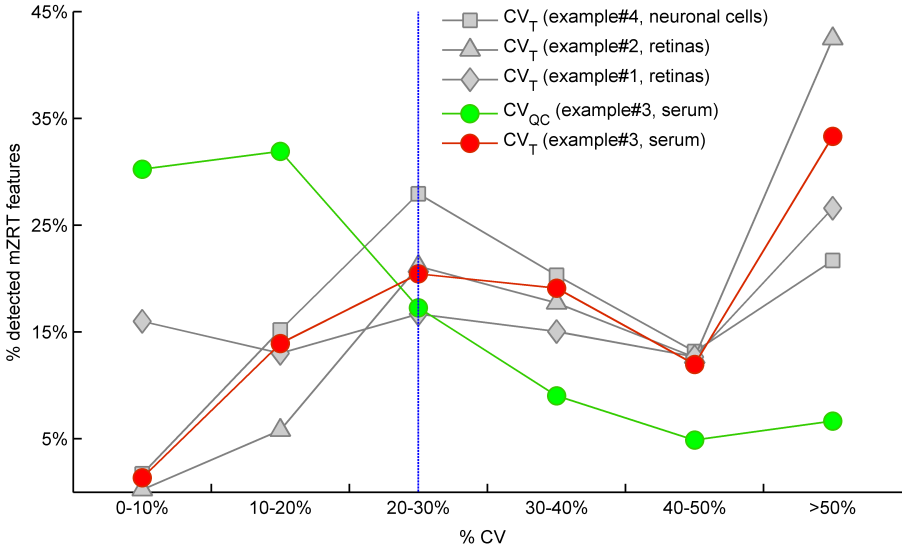


Figure 2. Comparison for our four working examples of the mzRT relative standard deviation (CV) frequency distributions calculated either across all the samples (CVT) or across QC samples (CVQC). Grey spots represent CVT for examples #1(), #2 (Δ) and #4 (☐) respectively. Green and red circles represent CVQC and CVT respectively for example #3. Blue line represents 20% CVT cut-off threshold established when QC samples are not available.

5. Hypothesis Testing

Untargeted metabolomics studies focused in this paper are aimed at the discovery of those metabolites that are varied between two

populations (*i.e.*, KO *vs* WT in examples #1 and 4 or treated *vs* untreated in example #3). In this sort of studies, random sample data from the populations to be compared are obtained in form of mzRT features dataset. Then, we calculate a statistic value (usually *mean* or *median*) and use statistical inference to determine whether the observed differences in the median or mean of the two populations are due to the phenomena under study or to randomness. Statistical inference is the process of drawing statements or conclusions about a populations based on sample data in a way that the risk of error of such conclusions is specified. These conclusions are based on probabilities arisen from evidences given by sample data [27].

To characterize those varied mzRT features, data sets are usually specified via hypothesis testing. Conventionally, we first postulate a null difference between the means/median of metabolic features detected in the populations under study by setting a *null hypothesis* (H0). Then, we specify the probability threshold for this null hypothesis to be rejected when in fact it is true. This threshold of probability called αis frequently set-up at 5% and it can be though as the probability of a false positive result or *Type I error*. Then, we use hypothesis testing to calculate the probability (*p-value*) of null hypothesis rejection. Whenever this p-value is bellow to this pre-defined threshold of probability (α), we reject the null hypothesis. On the other hand, when calculated p-values are larger than αwe do not have enough evidence to reject this hypothesis and we fail to reject it. Note that null hypothesis can never be proven, instead null hypothesis is either rejected or failed to reject. Conceptually, the failure to reject the null hypothesis (failure to find difference between

the means) does not directly translate in to accept or prove it (showing that there is no difference in reality).

A wide variety of univariate statistical tests to compare mean or medians are available. For a nonstatistician it can be daunting to figure out which one is most appropriate to implement with an untargeted metabolomic design and dataset. Helpful guidelines in basic statistics books can be consulted [27,28]. As summarized in Table 2, two important considerations should be taken in to account when deciding for a particular test. First one is the experimental design and second one data distribution.

Experimental design	**Normal distribution**	**Far from normal-curve**
	Compare Means	**Compare Medians**
Compare two unpaired groups	Unpaired t-test	Mann-Whitney
Compare two paired groups	Paired t-test	Wilcoxon signed-rank
Compare more than two unmatched groups	One-way ANOVA with multiple comparison	Kruskal-Wallis
Compare more than two matched groups	Repeated-measures ANOVA	Friedman

Table 2. Best suited statistical tests for datasets following normal distribution or far from the normal curve according to their experimental design

Experimental design will depend on experimental conditions considered when the metabolomics study is designed. Once the experimental design is fixed, population distribution determines the type of the test. Depending on this distribution, there are essentially two families of tests/ parametric and nonparametric. Parametric tests are based on the assumption that data are sampled from a Gaussian or normal distribution. Tests that do not make assumptions about the population distribution are referred as to non-parametric tests. Selection of parametric or non-parametric tests is not as clear-cut as might be a

priori though. Next section deals with the calculations necessary to guide such decision and exemplifies these calculations with our four working examples.

6. Deciding between Parametric or Non-Parametric Tests

6.1. Normality, Homogeneity of Variances and Independence Assumptions

Deciding between parametric and non-parametric tests should be based on three assumptions that should be checked/ normality, homogeneity of variances (*i.e.*, homocedasticity) and independence. Nevertheless, some of these assumptions rely on very theoretical mathematical constructs hardly ever met by real-life datasets obtained from metabolomics experiments.

Normality is assumed in parametric statistical tests such as t-test or ANOVA. Normal distributed populations are those presenting classical bell-shape curves to illustrate their probability density function. The frequency distribution of a normal population is a symmetric histogram with most of thefrequency counts bunched in the middle and equally likely positive and negative deviations from this central value. The frequencies of these deviations fall off quickly as we move further away from this central point corresponding to the mean. Data sampled from normal populations can be fully characterized by just two parameters/ the mean (x) and the standard deviation (σ). Normality assumption can be evaluated either statistically or graphically. We propose two tests to statistically evaluate normality/ Shapiro-Wilk and Kolmogorov-Smirnov, the former better behaved in the case of small samples sizes (*i.e.,* $N < 50$)

[27]. It is worth recalling that the term normal just applies to the entire population and not to the sample data. Hence, none of these tests would answer whether our dataset is normal or not. Their derived p-values must be interpreted as the probability of the data to be sampled from a normal distribution. On the other hand, testing normality is a matter of paradox/ for small samples sizes normality tests lack from power to detect non-normal distributions and as sample size increases normality becomes less troublesome thanks to the Central Limit Theorem. Since parametric tests are robust again mild violations of normality (and equality of variances as well), the practice of preliminary testing these two assumptions has been regarded as setting out in a rowing boat in order to test whether it is safe to launch an ocean liner [29]. Additionally, normality tests can be complemented with descriptive statistics such as Skewness and Kurtosis. On the other hand, graphical methods such as histograms, probability plots or Q-Q plots might result also helpful as tools to evaluate normality. Their use, however, is rather limited at exploratory stage of LC-MS untargeted metabolomic data since it is unfeasible to examine each one of these plots for each mzRT feature detected.

Another of the assumptions of a parametric test is that the within-group variances of the groups are all the same (exhibit homoscedasticity or homogeneity of variances). If the variances are different from each other (exhibit heteroscedasticity), the probability of obtaining a "significant" result even though the null hypothesis is true may be greater than the desired alpha level. There are both graphical and statistical methods for evaluating homoscedasticity. The graphical

method is the so-called boxplot but again, its use is rather limited because the impossibility to evaluate each one of them separately. The statistical methods are Levene's and Bartlett tests, the former the less sensitive to departures from normality. In both cases, the null hypothesis states that the group variances are equal. Resulting p-value < 0.05 indicate that the obtained differences in sample variances are unlikely to have occurred based on random sampling. Thus, the null hypothesis of equal variances is rejected and it is concluded that there is a difference between the population variances.

The third assumption refers to independence. Two events are independent when the occurrence of one event makes it neither more nor less probable that the other occurs. In our metabolomic context, the knowledge of the value of one sample entering the study provides no clue about the value of another sample to be drawn.

6.2. Parametric and Non-Parametric Tests. Does It Really Matters in LC-MS Untargeted Metabolomics Data?

Overall, the strength of violation of the three assumptions will determine the application of a parametric or non-parametric test. It should be noted that parametric tests are more powerful than non-parametric tests, *i.e.*, the use of a non-parametric test might miss a statistically significant difference that a parametric test would find. However, when dealing with non-normal populations, unequal variances, and unequal small sample sizes, a non-parametric test would perform better. This is the worst-case scenario for a parametric test to be non-robust. Although we recognize main weakness of normality testing, by

way of example we have calculated the percentage of features that meet normality and homocedasticity assumptions in our four working examples (Table 3).

	# mzRT	Groups	Normality (Shapiro-Wilk's test)	Homocedasticity (Levene's test)	Normality & Homocedasticity
Example #1 (Retinas)	3252	KO	66%	93%	60%
		WT	60%		54%
Example #2 (Retinas)	7654	Normoxia	65%	77%	48%
		Hypoxia	79%		60%
Example #3 (Serum)	6131	Untreated	85%	90%	76%
		Treated	88%		78%
Example #4 (Neuronal cells)	6831	KO	72%	91%	64%
		WT	82%		73%

Table 3. mzRT features percentages in which normality, homocedasticity or both assumptions are met. H0 (Shapiro-Wilk's test)= Data are sampled from a Gaussian distribution. H0 (Levene's test)=Variances are equal. Percentages represent those features in which there were not enough evidences to reject H0 at conventional α=0.05 relative to the total number of features retained after handling analytical variation.

According to Table 3 and considering the four examples on average, 65% of detected features meet normality and equality of variances assumptions. Therefore the use of a parametric test would be acceptable in 65% of the cases. Using a parametric test on the entire dataset would result in lack of robustness and consequent inaccurate p-values for the remaining 35% of features that do not meet parametric test assumptions. Alternatively, considering the use of a non-parametric would turn in loss of statistical power for those 65% of features. Alternatively we would transform those non-normally distributed data to normal or near to normal, for example taking logarithms when data come from a lognormal distribution. Nevertheless, data transformation

should be handled carefully since it might hamper the interpretation of the results.

To evaluate the consequences of using parametric or non-parametric tests in our datasets, we performed both types of tests and compare their outcomes. The Venn diagrams in Figure 3 show the percentage of features resulting in significantly different means/medians using parametric and non-parametric tests for the four working examples. Both tests share most of the significantly varying features and just a minor percentage of the total were specifically detected using either parametric or non-parametric tests. In general terms, analysis on the four working examples show a residual discrepancy between parametric or non-parametric test in terms of their outlined significant features. Although from these results we can not extrapolate a general methodology to choose between parametric and non-parametric tests, we recommend testing normality and equality of variances assumptions prior hypothesis testing to gain deeper insights in population distributions. Then, performing both parametric and non-parametric tests and to compare their outcomes prevailing parametric test outcomes for further calculations. Notice that if parametric and non-parametric tests result in high discrepancy we should check for outliers in our dataset.

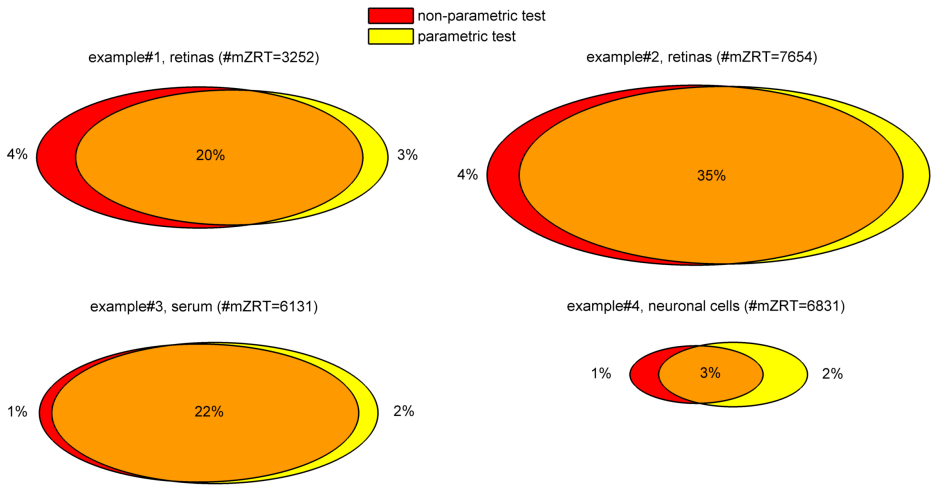


Figure 3. Venn-Diagrams of the mzRT features showing statistical significance using either parametric or non-parametric tests. Venn-Diagrams' areas are proportional to the percentage of the significantly varied features out of the number of total features retained after handling analytical variation (indicated in parenthesis) .The Mann-Whitney test (examples #1, 2 and 4) or Wilcoxon signed rank (example #3) tests were used for non-parametric groups median comparisons. Unpaired (examples #1, 2 and 4) or paired (example #3) t-tests were used for parametric groups mean comparisons.

7. Using Multiple Related Tests that Cumulate the p-Value/ The Multiple Testing Problem and the False Discovery Rate

7.1. The Multiple Testing Problem

In untargeted LC-MS-based metabolomics studies, the number of univariate-paralleled test equates to the number of mzRT features detected. As showed in our working examples, this number usually ranges in the thousands (it largely depends on experimental conditions). As the number of hypotheses tests increases, so as too does the probability of wrongly rejecting a null hypothesis because of random chance and therefore a substantial number of false positives (Type I error) might occur. This accumulation of false positives is termed the

multiple testing problem and is a general property of a confidence-based statistical test when applied across multiple features. From a metabolomics research standpoint, Type I errors are particularly undesirable. A substantial amount of work and resources based on MS/MS confirmation experiment can be stimulated in favor of a false finding. In the worstcase, a follow-up validation study on a false positive finding would not replicate the original work with consequent waste of resources and time. In such situations the chance for false positive rates must be carefully handed. Otherwise false findings may seriously affect the outcome of this type of studies [30]. Therefore, retrieved p-values from multiple tests performed in parallel across the detected mzRT features should be corrected. This is to re-calculate those probabilities obtained from a statistical test which is repeated multiple times. We are going to discuss two possible ways of handling multiple testing problem/ the Bonferroni and the FDR (False discovery Rate) corrections.

7.2. Bonferroni Correction

The family wise error (FWER) is defined as the probability of yielding one or more false positives out of all hypotheses tested. This error remains the most accepted parameter for ascribing significance levels to statistical test [31,32]. In multiple testing, if k independent comparison are performed FWER is increased at the rate of $1-(1-\alpha)k$; where k is the number of hypothesis tests performed and $\alpha$is the pre-defined threshold of probability in each individual test. Therefore, to maintain a prescribed FWER (*i.e.* 0.05) in an analysis involving multiple tests, the αassumed in each independent test must be more stringent than FWER. Bonferroni

correction is the standard approach to control FWER by specifying what αvalues should be considered in each individual test using the Equation 4

$$\alpha = FWDR/k \quad (4)$$

Considering our working example #1, 3252 mzRT features were retained after handling analytical variation. According to Bonferroni correction we should set a corrected $\alpha = 0.05/3252 = 1.054 \times 10^{-5}$ for each individual test to accept an overall FWER of 0.05. Hence, in each individual test, only those features with p-values $<1.54 \times 10^{-5}$ would be declared to be statistically significant. Assuming this correction, the probability of yielding one or more false positives out of all 3252 hypotheses tested would be $FWER = 1-(1-1.54 \times 10^{-5})3252 = 0.0488$. Notice that this probability is much lower than the one obtained if no correction was applied/ FWER = 1-(1-0.05)4581≈ 1. Bonferroni correction represents a substantial increase of the stringency of our testing leading to just 75 metabolite features out of the initially 3252 prescribing a FWER = 0.05.

Bonferroni correction keep a strict control on making one or more Type I error (false positive) at expenses of Type II errors (false negative). However, false negative findings might cause to overlook metabolites of potential interest and they also affect the outcomes of an untargeted metabolomics study. Other approaches to multiple testing correction such as the FDR (False Discovery Rate) claims for a striking balance between the concern about making too many false discoveries and the concern about missing the discovery of a real difference [33]. Next section deals on FDR correction and its interpretation.

7.3. The FDR Multiple Testing Correction

The FDR compute the number of false positives out of the significantly varied metabolic features, *i.e.*, the rate of significant features being false. This is different from the Bonferroni correction which focuses on the control on all falsely rejected hypotheses. In other fields such as microarray data experiments, the Bonferroni correction has been found to be too conservative and its use has led to many missed features of interest [33]. It has been argued that controlling the rate of allowed false findings using FDR do not represent a serious problem in the context of an exploratory research when further confirmatory studies are undertaken [31-33]. In addition, it has been demonstrated that controlling the FDR at the screening stage of the research carries a benefit for the next research stages[34]. Nevertheless, some authors in the field of metabolomics advocate that although being the most conservative, a Bonferroni analysis is both conceptually easier to understand and numerically easier to implement [35].

FDR correction calculates a p-corrected value or q-value for each tested metabolic feature. This qvalue is a function of the p-values and the distribution of the entire set of p-values from the family of tests being considered [31]. For each feature, its associated q-value can be though as the expected proportion of false positives considered when such feature is declared to be significantly varied. Hence, a metabolic feature having a q-value of 0.05 implies that 5% of metabolic features showing pvalues as small as such feature are false positives. A useful consideration is that a p-value of 0.05 implies that 5% of all tests will result in false positives and a q-value of 0.05 means that 5% out of the

significant tests will result in false positives. A useful plot to evaluate the proportions of false positives is a frequency histogram illustrating the

distribution of p-values obtained from paralleled tests across all mzRT features in a dataset. Figure 4 illustrates such histograms for examples #1, #2 and #4. Those mzRT features with significant changes in their relative abundance will show small p-values and therefore the histogram will be skewed towards 0 (examples #1 and 2). On the contrary, metabolic features showing no change in their relative abundances will show a uniform random flatten frequency distribution (example #4). The green bar represents those metabolic features declared to be significant in the t-test binary group comparison for each example ($p < 0.05$). The actual FDR calculated proportion of such features resulting in false positives correspond to the red bar (q-values > 0.05).

According to Figure 4, t-test comparison of KO and WT groups in example#1 lead to 708 significantly changed metabolic features out of 3252. By setting our αthreshold to 5% we accepted 163 features to be false positives. This represents 23% out of the 708 features significantly varied. Notice that after FDR correction we obtained 453 mzRT features with q-values bellow 5% of false positives acceptance threshold. This means that 5% out of this 453 mzRT features (*i.e.*, 23) are expected to be false positives. An acceptance of 5% chance of false positives results in a better situation than the one derived if no correction was applied (meaning 23% chance of false positives). Recall that in this same example, Bonferroni correction lead to consider just 75 features with an adjusted threshold p-value< 1.54×10^{-5}. Bonferroni provides the strongest

control of the false positives and therefore a high confidence in the selected metabolic features. However, an important advantage of FDR approach is that it allows the researcher to select the error rate that they would assume in their subsequent studies. On the other hand, Figure 4 show that a t-test comparison of WT and KO groups on example#4 outlined 328 features all of them resulting in false positives after FDR correction. This indicates that all this significant outcomes derived from chance and no real effect was underlying on this example. Accordingly if no correction for multiple testing were considered we would have done

subsequent MS/MS identification experiments on features that represent false positives. This would have been a pointless task with consequent waste of time and resources. To avoid situations like this, we would recommend correcting for multiple testing when dealing with multiple univariate analysis of untargeted LC-MS datasets. Then, focus on those metabolites with lower FDR derived q-values for further MS/MS identification experiments. In addition, we would like to comment that whenever a follow-up targeted validation study was going to be attempted, we would recommend considering those metabolites showing statistical significance after strict Bonferroni correction.

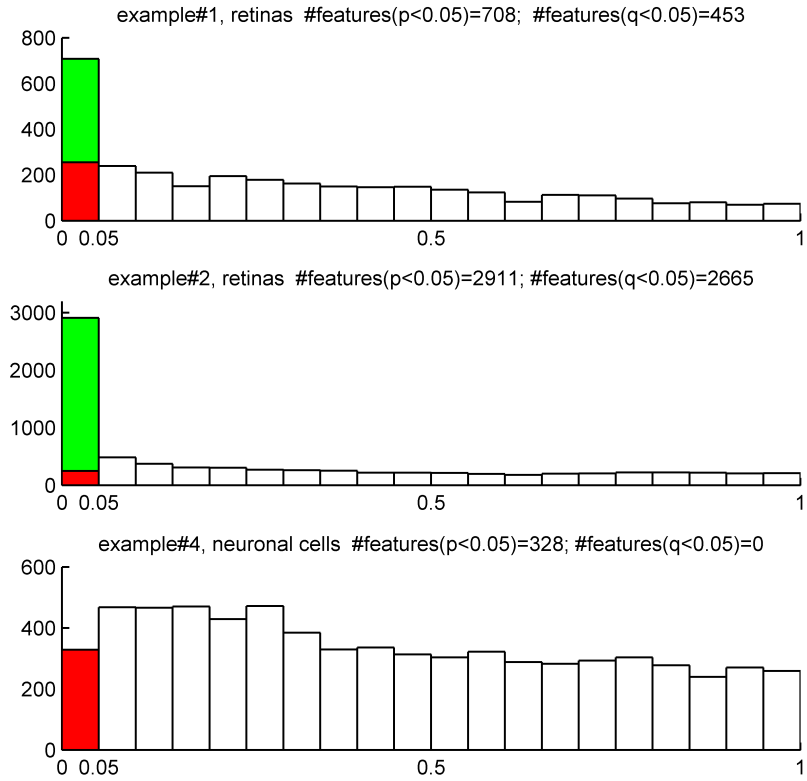


Figure 4. Frequency histogram showing the distribution of p-values typically expected from t-tests binary groups' comparison in examples #1, 2 and 4. Green bar represent the total number of features declared to be significant assuming 5% false positives in a t-test comparison of the two groups. Red bar represent the FDR- estimated number of features being considered false positives out of the features declared significant in the t-test. The number of total significant features retained after FDR correction ($q < 0.05$) is also indicated.

8. The Fold Change Criteria

A common practice to identify mzRT features of relevance within a dataset is to rank these features according their fold change (FC). FC can be though as the magnitude of difference between the two populations under study. For each mzRT feature, a FC value is computed according to equation 5 in which X represents the average raw intensities across

"case" group and Y represents the average raw intensities across "control" group. Whenever the raw intensities of the "control" group are larger than in the "case" group, this ratio should be inverted and sign should be conveniently changed to indicate a decrease of the case group relative to the control. Of mention, in paired-data designs, fold change should be calculated as the average of each individual fold change across all sample pairs.

$$FC_{mZRT} = \frac{\bar{X}}{\bar{Y}}, \quad \bar{X} > \bar{Y}; \; FC_{mZRT} = -\frac{\bar{Y}}{\bar{X}} \quad \bar{X} < \bar{Y} \tag{5}$$

In formal statistical terms, a mzRT feature is claim to be varied among two conditions when its relative intensity values change systematically between these two condition regardless on how small this change is. However, significance does not contain information about the magnitude of this change. For a metabolomics standpoint, a metabolic feature is considered to be relevant only when this change result in a worthwhile amount. Hence, significantly varied mZRT are ranked according to their FC value. Subsequent MS/MS chemical structural identification experiments are performed on those metabolic features resulting above a minimum FC cutoff value. It has been demonstrated that a 2-FC cutoff for metabolomics studies using human plasma or CSF minimizes the effects of biological variation inherent in a healthy control group [26]. However, this cutoff value is set rather arbitrarily and based on similar FC cutoff values routinely applied in gene chip experiments.

9. Univariate LC-MS Untargeted Analysis Workflow

The typical univariate data analysis flow diagram for untargeted LC-MS metabolomics experiments is summarized in Figure 5. The ultimate goal is to constraint the number of initially detected mzRT features to an amenable number for further MS/MS identification experiments. Only those mZRT features showing both statistically significant changes with delimited chance for false positives in their relative intensity and a minimum FC are going to be retained. Steps 1-5 are below summarized/

STEP1. Use quality control check to get rid-out of those mZRT features that do not contain biological information. Ideally QC samples should be measured. Then, compute CVQC and proceed to retain only those metabolic features presenting CVQC < 20%. If QC samples are not available, an alternative procedure is to compute CVT and retain those mZRT with CVT > 20%.

STEP2. Mind the experimental design to select the best suited statistical test to apply. Check whether your data is paired or not, *i.e.*, whether your groups are related such as in our example#3 (individuals prior to treatment are uniquely matched to the same individual after the treatment). Afterwards, check normality and equality of variances assumptions. Be aware that performances of the normality tests might be hampered by low samples sizes dataset commonly found in LC-MS untargeted metabolomics studies. Despite this, working on such tests might be useful to gain some insights in to the data distribution.

STEP 3. Compare mean or medians of your dataset performing statistical inference and trying to apply statistical tests thoughtfully instead of mechanically. Try to be aware of the tests weaknesses when

applying it. Once we have taken the decision on whether using parametric or non-parametric tests, it is important to stick on the same approach through the rest of the data analysis procedure. Thisis to plot our results in the form of medians instead of means whenever we choose to use a nonparametric statistical test.

STEP4. Account for multiple testing. Report the number of positive false findings after FDR correction. Plot histograms of p-values frequency distribution to get an overview of whether a dataset contains significant differences. Decide a FDR threshold to accept. A general consensus is to accept 5% of FDR level but there is nothing special about this value and each researcher might justify their assumed FDR value, which should be fixed before data is collected.

STEP5. Compute mean or median FC depending on the test used to perform statistical inference. Fix a cutoff FC value. From our in-house experience we recommend an arbitrary 1.5-FC cutoff valuemeaning a minimum of 50% of variation in the two groups compared. Rank your significant list of features according the FC value. Retain those significant features with higher FC values for MS/MS experiments and follow-up validation studies.

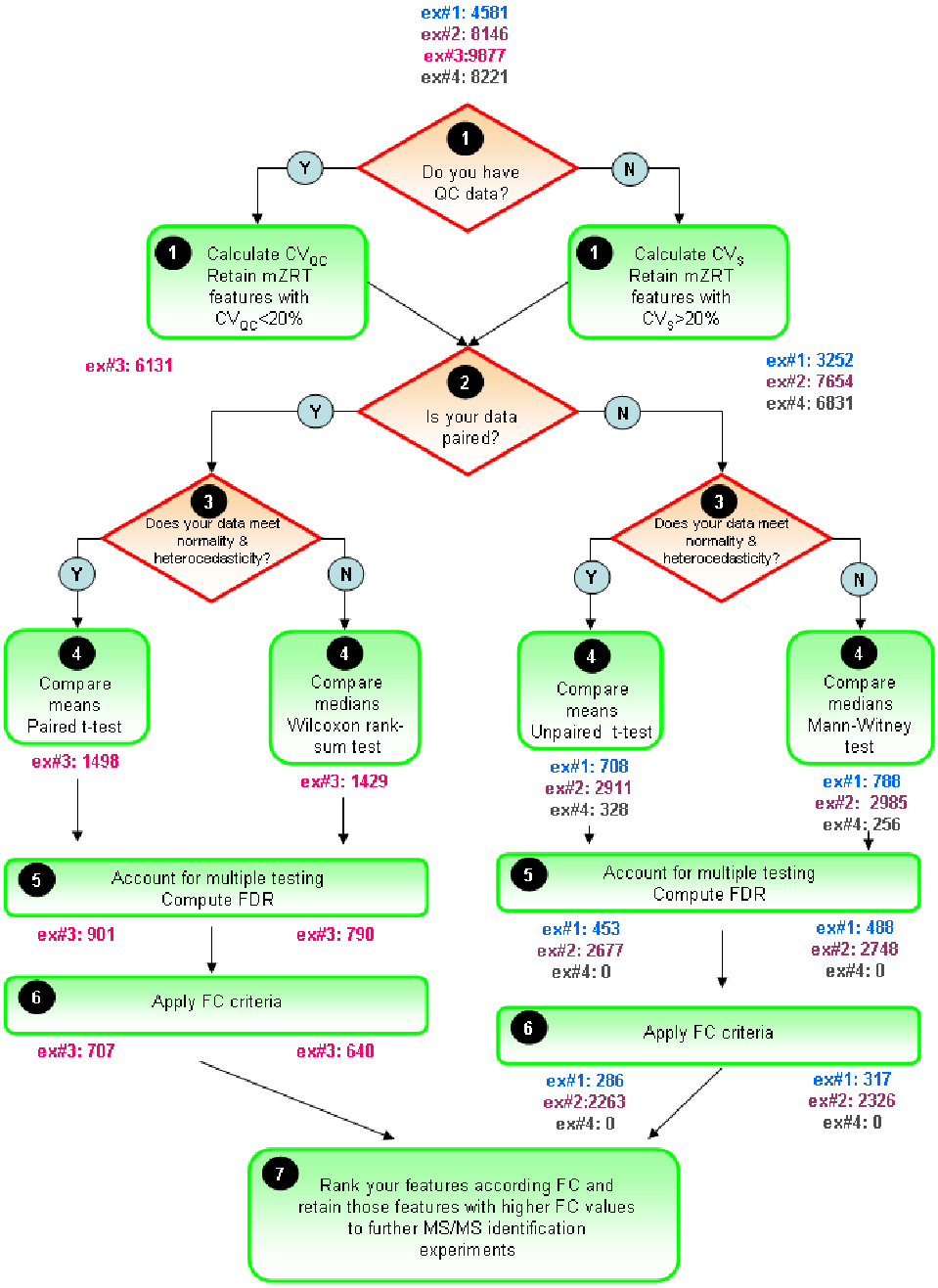


Figure 5. General flow chart for univariate data analysis of untargeted LC-MS-based metabolomics data. Different colors for the four working examples indicate the initial number and the retained number of mzRT features in each step. FDR and FC value are fixed at 5% level 1.5-cutoff values respectively.

Following steps 1-5 described above, those metabolites identified using MS/MS experiments for example #2 are summarized in Table 4. Of mention all metabolites identified meet the statistical criteria described above regardless of using either parametric or non-parametric tests. Notice the small number of properly identified metabolites as compared to the high number of features surviving statistical criteria. It is important to mention that in the best optimistic case the number of metabolite identifications showing MS/MS confirmation use to be in the tens after a formal untargeted metabolomics experiment. Conversely, in case of putative identifications based on exact masses, the number of metabolites reported is much higher. However, recall that such metabolites are just putatively identified. Considering that replication experiments are necessary to undeniably ascertain the role of the metabolites found to be relevant in the untargeted study, a strict identification of the metabolites is essential. In this sense, our work-flow data analysis represents the first step for a successful identification of those metabolites.

	Parametric Test			Non-parametric		
	p-value	q- value	FC (mean)	p-value	q-value	FC (median)
Hexadecenoylcarnitine	3.31×10^{-13}	1.05×10^{-10}	5.0	2.49×10^{-05}	3.18×10^{-04}	**4.9**
Acetylcarnitine-derivative	1.10×10^{-13}	5.02×10^{-11}	7.2	2.49×10^{-05}	3.18×10^{-04}	**7.5**
Tetradecenoylcarnitine	1.29×10^{-13}	5.29×10^{-11}	8.8	2.49×10^{-05}	3.18×10^{-04}	**8.8**
Decanoylcarnitine	7.79×10^{-11}	1.03×10^{-08}	5.7	2.49×10^{-05}	3.18×10^{-04}	**5.6**
Laurylcarnitine	8.48×10^{-11}	1.06×10^{-08}	9.2	2.49×10^{-05}	3.18×10^{-04}	**8.7**
7-ketocholesterol	4.00×10^{-09}	1.92×10^{-07}	3.1	2.49×10^{-05}	3.18×10^{-04}	**3.3**
5,6β-epoxy-cholesterol	2.12×10^{-08}	6.61×10^{-07}	5.1	2.49×10^{-05}	3.18×10^{-04}	**7.0**
7α-hydroxycholesterol	3.88×10^{-08}	1.07×10^{-06}	4.1	2.49×10^{-05}	3.18×10^{-04}	**4.5**
All-trans-Retinal	1.26×10^{-05}	9.24×10^{-05}	-3.0	4.01×10^{-05}	3.98×10^{-04}	**-2.8**
Octanoylcarnitine	$\mathbf{9.21\times10^{-05}}$	$\mathbf{4.28\times10^{-04}}$	**5.5**	$\mathbf{5.09\times10^{-03}}$	$\mathbf{1.14\times10^{-02}}$	**17.2**

Table 4. Statistics summary of those metabolites identified using MS/MS experiments in working example #2. Unpaired t-test and Mann-Whitney test were used for parametric and non-parametric hypoxic and normoxic retinas comparison respectively. Correction for multiple testing was performed assuming 5% FDR.

REFERENCES AND NOTES

1. Patti, G.J.; Yanes, O.; Siuzdak, G. Innovation/ Metabolomics/ the apogee of the omics trilogy. *Nat. Rev. Mol. Cell. Biol.* 2012, *13*, 263–269.

2. Smith, C.A.; Want, E.J.; O'Maille, G.; Abagyan, R.; Siuzdak, G. XCMS/ Processing Mass Spectrometry Data for Metabolite Profiling Using Nonlinear Peak Alignment, Matching, and Identification. *Anal. Chem.* 2006, *78*, 779–787.

3. Katajamaa, M.; Oresic, M. Processing methods for differential analysis of LC/MS profile data. *BMC Bioinformatics* 2005, *6*, 179.

4. Lommen, A. MetAlign/ Interface-Driven, Versatile Metabolomics Tool for Hyphenated Full-Scan Mass Spectrometry Data Preprocessing. *Anal. Chem.* 2009, *81*, 3079–3086.

5. Kuhl, C.; Tautenhahn, R.; Böttcher, C.; Larson, T.R.; Neumann, S. CAMERA/ An Integrated Strategy for Compound Spectra Extraction and Annotation of Liquid Chromatography/Mass Spectrometry Data Sets. *Anal. Chem.* 2011, *84*, 283–289.

6. Alonso, A.; Julia, A.; Beltran, A.; Vinaixa, M.; Diaz, M.; Ibanez, L.; Correig, X.; Marsal, S. AStream/ an R package for annotating LC/MS metabolomic data. *Bioinformatics* 2011, *27*, 1339–1340.

7. Kristian Hovde, L. Multivariate methods in metabolomics – from pre-processing to dimension reduction and statistical analysis. *Trac-Trend. Anal. Chem.* 2011, *30*, 827–841.

8. Hendriks, M.M.W.B.; Eeuwijk, F.A.v.; Jellema, R.H.; Westerhuis, J.A.; Reijmers, T.H.; Hoefsloot, H.C.J.; Smilde, A.K. Data-processing strategies for metabolomics studies. *Trac-Trend. Anal. Chem.* 2011, *30*, 1685–1698.

9. Kalogeropoulou, A. Pre-processing and analysis of high-dimensional plant metabolomics data. Master Thesis, University of East Anglia, Norwich, UK, 2011.

10. Goodacre, R.; Broadhurst, D.; Smilde, A.; Kristal, B.; Baker, J.; Beger, R.; Bessant, C.; Connor, S.; Capuani, G.; Craig, A.; et al. Proposed minimum reporting standards for data analysis in metabolomics. *Metabolomics* 2007, *3*, 231–241.

11. Karp, N.A.; Griffin, J.L.; Lilley, K.S. Application of partial least squares discriminant analysis to two-dimensional difference gel studies in expression proteomics. *Proteomics* 2005, *5*, 81–90.

12. Kenny, L.C.; Broadhurst, D.I.; Dunn, W.; Brown, M.; North, R.A.; McCowan, L.; Roberts, C.; Cooper, G.J.S.; Kell, D.B.; Baker, P.N.; et al.

Robust Early Pregnancy Prediction of Later Preeclampsia Using Metabolomic Biomarkers. *Hypertension* 2010, *56*, 741–749.

13. R Development Core Team. 2009 R/ A language and environment for statistical computing. Available online/ http///www.R-project.org, accessed on 17 October 2012.

14. Patti, G.J.; Yanes, O.; Shriver, L.P.; Courade, J.P.; Tautenhahn, R.; Manchester, M.; Siuzdak, G. Metabolomics implicates altered sphingolipids in chronic pain of neuropathic origin. *Nat. Chem. Biol.* 2012, *8*, 232–234.

15. Yanes, O.; Clark, J.; Wong, D.M.; Patti, G.J.; Sanchez-Ruiz, A.; Benton, H.P.; Trauger, S.A.; Desponts, C.; Ding, S.; Siuzdak, G. Metabolic oxidation regulates embryonic stem cell differentiation. *Nat. Chem. Biol.* 2010, *6*, 411–417.

16. Marchetti, V.; Yanes, O.; Aguilar, E.; Wang, M.; Friedlander, D.; Moreno, S.; Storm, K.; Zhan, M.; Naccache, S.; Nemerow, G.; et al. Differential macrophage polarization promotes tissue remodeling and repair in a model of ischemic retinopathy. *Sci. Rep.* 2011, *1*, 76.

17. Dorrell, M.I.; Aguilar, E.; Jacobson, R.; Yanes, O.; Gariano, R.; Heckenlively, J.; Banin, E.; Ramirez, G.A.; Gasmi, M.; Bird, A.; et al. Antioxidant or neurotrophic factor treatment preserves function in a mouse model of neovascularization-associated oxidative stress. *J. Clin. Invest.* 2009, *119*, 611–623.

18. Vinaixa, M.; Rodriguez, M.A.; Samino, S.; Díaz, M.; Beltran, A.; Mallol, R.; Bladé, C.; Ibañez, L.; Correig, X.; Yanes, O. Metabolomics Reveals Reduction of Metabolic Oxidation in Women with Polycystic

Ovary Syndrome after Pioglitazone-Flutamide-Metformin Polytherapy. *PloS One* 2011, *6*, e29052.

19. Grainger, D.J. Megavariate Statistics meets High Data-density Analytical Methods/The Future of Medical Diagnostics? *IRTL Rev. 1* 2003, 1–6.

20. Ferreira, J.A.; Zwinderman, A. Approximate sample size calculations with microarray data/ an illustration. *Sta.t Appl. Genet. Mol. Biol.* 2006, *5*, Article25.

21. Ferreira, J.A.; Zwinderman, A.H. Approximate Power and Sample Size Calculations with the Benjamini-Hochberg Method. *Int. J. Biostat.* 2006, *2*.

22. van Iterson, M.; 't Hoen, P.; Pedotti, P.; Hooiveld, G.; den Dunnen, J.; van Ommen, G.; Boer, J.; Menezes, R. Relative power and sample size analysis on gene expression profiling data. *BMC Genomics* 2009, *10*, 439.

23. van der Kloet, F.M.; Bobeldijk, I.; Verheij, E.R.; Jellema, R.H. Analytical Error Reduction Using Single Point Calibration for Accurate and Precise Metabolomic Phenotyping. *J. Proteome. Res.* 2009, *8*, 5132–5141.

24. Dunn, W.B.; Broadhurst, D.; Begley, P.; Zelena, E.; Francis-McIntyre, S.; Anderson, N.; Brown, M.; Knowles, J.D.; Halsall, A.; Haselden, J.N.; *et al.* Procedures for large-scale metabolic profiling of serum and plasma using gas chromatography and liquid chromatography coupled to mass spectrometry. *Nat. Prot.* 2011, *6*, 1060–1083.

25. Masson, P.; Alves, A.C.; Ebbels, T.M.D.; Nicholson, J.K.; Want, E.J. Optimization and Evaluation of Metabolite Extraction Protocols for

Untargeted Metabolic Profiling of Liver Samples by UPLC-MS. *Anal. Chem.* 2010, *82*, 7779–7786.

26. Crews, B.; Wikoff, W.R.; Patti, G.J.; Woo, H.-K.; Kalisiak, E.; Heideker, J.; Siuzdak, G. Variability Analysis of Human Plasma and Cerebral Spinal Fluid Reveals Statistical Significance of Changes in Mass Spectrometry-Based Metabolomics Data. *Anal. Chem.* 2009, *81*, 8538-8544.

27. Riffenburgh, R.H. *Statistics in Medicine*; Elsevier/ Amsterdam, The Netherland, 2006. 28. Motulsky, H. *Intuitive Biostatistics*; Oxford University Press/ New York, NY, USA, 1995.

29. Box, G.E.P. Non-Normality and Tests on Variances. *Biometrika* 1953, *40*, 318–335.

30. Ioannidis, J.P.A. Why Most Published Research Findings Are False. *PLoS Med.* 2005, *2*, e124.

31. Storey, J.D.; Tibshirani, R. Statistical significance for genomewide studies. *Proc. Natl. Acad. Sci. USA* 2003, *100*, 9440–9445.

32. Storey, J.D. A direct approach to false discovery rates.*J. Roy. Stat. Soc. B Met.* 2002, *64*, 479–498.

33. Benjamini, Y.; Drai, D.; Elmer, G.; Kafkafi, N.; Golani, I. Controlling the false discovery rate in behavior genetics research. *Behav. Brain. Res.* 2001, *125*, 279–284.

34. Benjamini, Y.; Yekutieli, D. Quantitative Trait Loci Analysis Using the False Discovery Rate. *Genetics* 2005, *171*, 783–790.

35. Broadhurst, D.; Kell, D. Statistical strategies for avoiding false discoveries in metabolomics and related experiments. *Metabolomics* 2006, *2*, 171–196.

2.2 Clinical Metabolomics: PCOS

2.2.1 New metabolic insights in PCOS

2.2.1.1 Metabolic heterogeneity in polycystic ovary syndrome (PCOS) is determined mainly by obesity: a plasma untargeted gas chromatography-mass spectrometry (GC-MS) metabolomic approach

SUMMARY

BACKGROUND: Abdominal adiposity and obesity influence the association of polycystic ovary syndrome (PCOS) with insulin resistance and diabetes. We aimed to characterize the intermediate metabolism phenotypes associated with PCOS and obesity.

METHODS: We applied a untargeted gas chromatography-mass spectrometry (GC-MS) metabolomic approach to plasma samples from 36 patients with PCOS and 39 control women without androgen excess, matched for age, body mass index and frequency of obesity.

RESULTS: Patients with PCOS were hyperinsulinemic and insulin resistant compared with the controls. The increase in plasma long-chain fatty acids such as linoleic and oleic acid and of glycerol found in the obese patients with PCOS suggests increased lipolysis, possibly secondary to impaired insulin action at adipose tissue. On the contrary, non-obese patients with PCOS showed a metabolic profile consisting of suppression of lipolysis and increased glucose utilization (increased lactic acid levels) at peripheral tissues, and PCOS patients as a whole showed decreased 2-ketoisocaproic and alanine levels suggesting utilization of branch-chain aminoacids for protein synthesis and not for gluconeogenesis. These metabolic processes required effective insulin signaling; therefore, insulin resistance was not universal in all tissues of

these women, and different mechanisms possibly contributed to their hyperinsulinemia. PCOS was also associated with decreased α-tocopherol and cholesterol levels irrespective of obesity.

CONCLUSIONS: Substantial metabolic heterogeneity, strongly influenced by obesity, underlies PCOS. The possibility that hyperinsulinemia may occur in the absence of universal insulin resistance in non-obese women with PCOS should be considered when designing diagnostic and therapeutic strategies for the management of this prevalent disorder.

INTRODUCTION

PCOS is the most prevalent endocrine disorder in women of fertile age [83]. PCOS is characterized by androgen excess and is associated with metabolic disorders such as obesity, insulin resistance and diabetes [84,85,86,87]. The occurrence of disorders of glucose tolerance and diabetes in women with PCOS appears to be mainly dependent on the association with insulin resistance and obesity playing a major triggering role [82]. Whether the increased risk of diabetes pertains to all patients of PCOS irrespective of obesity, or only to obese patients and non-obese patients with additional risk factors, is still matter of debate [88,89].

The application of hypothesis-free genomic and proteomic techniques to the study of PCOS confirmed the involvement of insulin resistance and low-grade chronic inflammation in its pathogenesis [90], but also revealed the participation of metabolic pathways related to iron metabolism, Wnt signaling, oxidative stress, immune function, and lipid metabolism in the pathogenesis of this prevalent disorder [75,90,91,92].

Recently introduced metabolomics aim at the identification and quantification of all metabolites in biological systems [18]. Metabolites include a wide array of compound classes such as amino acids, lipids, organic acids, nucleotides and others which are quite diverse in their physical and chemical composition and occur in a wide range of concentrations [18]. Compared with genomics, transcriptomics and proteomics, metabolomics provides data that are the most predictive of the phenotype [18] because the metabolome is the net result of genomic, transcriptomic and proteomic variability, and of their interaction with the environment [93].

Untargeted hypothesis-free metabolomic approaches intend to compare as many metabolites as possible of different samples, reporting those metabolites that quantitatively differ based on statistical analysis [93]. As other *omics* techniques, untargeted metabolomic profiling is not constricted by previous knowledge on the field and could indicate novel mechanisms of disease and even therapeutic targets that may have been lost by classic targeted approaches.

In order to provide new insights on the impact that obesity exerts on the metabolic derangements frequently associated with PCOS, we have characterized the intermediate metabolism phenotypes associated with PCOS and obesity in premenopausal women by applying a untargeted gas chromatography-mass spectrometry (GC-MS) based metabolomic approach to their plasma samples.

MATERIALS AND METHODS

Subjects

We selected plasma samples from 36 patients with PCOS and 39 control women who had no evidence of androgen excess or ovulatory dysfunction. We selected both groups from our historical database of patients and controls using age and body mass index (BMI) as sole criteria, in order to avoid any selection bias related to prior knowledge of their metabolic or hormonal characteristics. The selection process aimed to have similar age, BMI and frequencies of obesity, as defined by a BMI ≥ 30 kg/m^2, in both groups.

The diagnosis of PCOS relied on the presence of clinical and/or biochemical, hyperandrogenism, oligoovulation, and exclusion of secondary etiologies as previously reported [94]. Evidence of oligoovulation was provided by serum luteal phase progesterone measurements less than 4 ng/ml (12.7 pmol/l) in patients with regular menses, by oligomenorrhea (more than six cycles longer than 36 d in the previous year) or by amenorrhea (absence of menstruation for 3 consecutive months). Although ovarian morphology was not analyzed, by having hyperandrogenism and oligo-ovulation all patients fulfilled all the current definitions of PCOS [78,79,95], and PCOS was ruled out reliably in the controls because all these women did not have hyperandrogenism and showed regular ovulatory menstrual cycles.

None of the women had either a personal history of hypertension, diabetes mellitus or cardiovascular events, or had received treatment with oral contraceptives, antiandrogens, insulin sensitizers, iron supplements or drugs that might interfere with blood pressure

regulation, lipid profile or carbohydrate metabolism for the previous 6 months. We obtained written informed consent from all the participants, and approval from the local Ethics Committee.

We obtained plasma samples during the follicular phase of the menstrual cycle or during amenorrhea, after a 3-day 300-g carbohydrate diet and 12-h overnight fasting. The technical characteristics of the assays employed for plasma glucose and lipids and serum hormone measurements have been described elsewhere [94,96]. The composite insulin sensitivity index [97] was determined from the glucose and insulin measurements obtained during a standard oral glucose tolerance test.

Plasma handling

GC-MS analysis was performed according to Agilent's specifications [98]. One-hundredµL plasma aliquots were thawed to 4°C. After thawing, samples were briefly vortexed to remove inhomogeneities due to the freezing-thawing process. Each aliquot was spiked with 20 µl internal standard solution (1 µg/µl succinic-d4 acid; Sigma-Aldrich Química SA, Madrid, Spain). Then proteins were precipitated by the addition of 900 µl of cold methanol/water (8/1 v/v) followed by 4 minutes ultrasonication and 10 seconds vortexing. Aliquots were kept subsequently on ice for 10 minutes. After centrifugation (10 minutes at 19,000 g at 4ºC), two technical replicates of 200 µl each of the supernatant was transferred to a GC autosampler vial and 20 µl of myristic-d27 acid (Sigma Aldrich) used as internal standard for retention time lock (RTL system provided in Agilent's ChemStation Software, Agilent Technologies Inc., Santa Clara, CA) was added to each aliquot. Finally, the mixture was lyophilized

overnight using a Lyotrap freeze-dryer (LTE Scientific Ltd, Greenfield, United Kingdom).

The first step of the derivatization was methoximation to prevent ring formation and to stabilize carbonyl moieties. Thus, we incubated lyophilized plasma residues with 50 µL of methoxyamine in pyridine (0.3 µg/µl) for 16 hours at room temperature. To increase the volatility of the compounds, in the next step we silylated the samples using 30 µl of N-methyl-N-trimethylsilyltrifluoroacetamide with 1%trimethylchlorosilane (MSTFA + 1% TMCS, Thermo Fisher Scientific Inc., Alcobendas, Spain) for 1 hour at room temperature.

GC-MS analysis

We submitted derivatized plasma samples to a GC-MS system that consisted of a HP6890 gas chromatograph coupled to a quadrupole HP5973 mass spectrometric detector (Agilent Technologies Inc.). The detector had a resolving power of unit mass resolution over mass range m/z 5 to 500 and mass accuracy of ± 0.2u. The system was equipped with electron impact ionization (EI), an autosampler (7683 series), an injector (7683B series) and the ChemStation software (G1712DA, Rev. 02.00). We used the retention time locking software (RTL, Agilent Technologies Inc.) to stabilize retention times from run to run.

The head pressure was adjusted in the constant pressure mode to the locked retention time 16.727 min (17.5 psi) of the reference compound myristic-d27 acid. One µL of each derivatized sample was injected in the gas chromatograph system with a split inlet (split ratio 5/1) equipped with a J&W Scientific DB 5-MS+DG stationary phase column (30 mm × 0.25 mm i.d., 0.1 µm film, Agilent Technologies Inc.)

coupled to the HP5973 mass selective detector. The column temperature was held at 60 ºC for 1 min, then increased to 325ºC at a rate of 10 ºC/ min and finally held at this temperature for 10 min. The detector was operated in the scan mode from the electron impact ionization mode (70 eV). Mass spectra were recorded after a solvent delay of 4 min with 2.46 scans per second (mass scanning range of m/z 50–600; threshold abundance value of 50 counts). The source temperature and quadrupole temperature were 230 and 150 °C, respectively.

We used quality control samples (QCs) consisting of pooled plasma samples to check the GC-MS system for reproducibility. QCs were injected periodically after 12-study samples. The reproducibility of our QCs peak areas in terms of % RSD was 29% for cholesterol, 6% for glucose and 14% for glycerol, which are within the acceptable range of less than 30% RSD [99]. To check the reproducibility of our ionization efficiency in plasma, we performed 5 repeated GC-MS injection of the same plasma sample spiked at a final concentration of 3mg/mL of myristic acid d27. We obtained a 5% RSD on the area of its representative ion (m/z 312 at RT=16.727 min). Moreover, injecting 6 times another plasma sample provided a 15% RSD for a set of 10 known metabolites spanning the whole chromatogram. In addition, to begin with the chromatographic analysis, injections of 3 blank runs were performed daily to check for carry-over effects. The experimental samples were carefully randomized to ensure that potential experimental drifts affected the experimental groups to the same extent.

GC-MS data analysis

Raw GC-MS files were exported in the platform-independent netCDF (*.cdf) and loaded in to XCMS software (version 1.6.1) based on R-program version 2.9.2 (R-Foundation for statistical computing, www.Rproject.org). After automatic peak peaking, integration of chromatographic peaks and alignment in the time domain XCMS feature table containing integrated intensities of each mass-to-charge ratio retention time (m/zRT) pair in a tab-separated text file (*.txt). We then used Matlab (TheMathWorks, Inc., Natick, Massachusetts) to extract and normalize to internal standard succinic-d4 acid features that were present in both plasma aliquots of 80% or more of the samples of at least one of the experimental groups. Subsequently, we performed statistical analysis (see below) of the normalized mZRT features obtained from XCMS. Only those fragments showing statistically significant differences among the groups of women were used to annotate peaks. Groups of features sharing the same retention time that turned out to be statistically significant and that presented with a high degree of correlation were considered representative of a single metabolite. For metabolite annotation, AMDIS program (Automated Mass Spectral Deconvolution and Identification System, National Institute of Standards and Technology, Gathersburg, MD) was run for peak deconvolution and both the Fiehn GC-MS Metabolomics RTL Library and NIST mass spectral databases were used for tentative identification (match with library > 65%).

Statistical Analysis

The influence of PCOS status and obesity and the interaction between both independent variables on the m/zRT intensities obtained by XCMS software, normalized to internal standard succinic acid-d4,were analyzed by two-way ANOVA. Clinical variables were analyzed after confirming normality of dependent variables by the Kolmogorov-Smirnov test. GC-MS variables were rank transformed before submitting them to two-way ANOVA, as most of them did not follow a normal distribution. Because two-way ANOVA incorporates multiple comparisons among groups into its calculations, we set the level of statistical significance at the $\alpha = 0.05$ level.

RESULTS

As expected from design, we found no differences in age and body mass index among patients with PCOS and controls. However, when compared with the controls, patients with PCOS presented with increased waist hip ratio, hirsutism score, free testosterone, dehydroepiandrosterone-sulfate and fasting insulin levels, and decreased fasting glucose and insulin sensitivity index values (Table 1), irrespective of the occurrence of obesity.

When considered as a whole, obese women presented with increased body mass index and waist hip ratio, free testosterone concentrations, fasting insulin, LDL-cholesterol levels and triglycerides, and reduced sex hormone-binding globulin and HDL-cholesterol levels and decreased insulin sensitivity index compared with their non-obese counterparts (Table 1). We did not find any statistically significant

interaction between PCOS status and obesity, indicating that the effects of PCOS were common to non-obese and obese women and, *viceversa*, the effects of obesity were the same in the patients with PCOS and in the controls (Table 1).

	P-value		
	PCOS	Obesity	Interaction
Age (yr)	0.871	0.089	0.853
Body mass index (kg/m^2)[a]	0.643	<0.001	0.933
Waist hip ratio[a,b]	0.023	<0.001	0.258
Frequency of smokers	0.817	0.351	
Hirsutism score[b]	<0.001	0.755	0.154
Free testosterone (ng/dl)[a,b]	<0.001	<0.001	0.228
Sex hormone-binding globulin (µg/dl)[a]	0.094	<0.001	0.796
Dehydroepiandrosteron e-sulfate (ng/ml)[b]	0.012	0.454	0.878
Fasting glucose (mg/dl)[b]	0.022	0.225	0.886
Fasting insulin (µU/ml)[a,b]	0.008	<0.001	0.714
Insulin sensitivity index[a,b]	0.022	0.001	0.263
Total cholesterol (mg/dl)	0.793	0.126	0.563
HDL-cholesterol (mg/dl)[a]	0.994	0.010	0.063
LDL-cholesterol (mg/dl)[a]	0.487	0.025	0.835
Triglycerides (mg/dl)[a]	0.070	<0.001	0.243

Table 1. Clinical and hormonal characteristics of patients with PCOS compared with control women. Differences in continuous variables among groups were analyzed by 2-way ANOVA, and categorical data were analyzed by X^2tests.

Table 2 summarizes detailed peak annotation, assignment and univariate statistical analysis derived from the study of the interplay between PCOS and obesity. 2-Ketoisocaproic acid, α-tocopherol, alanine and cholesterol were different in patients with PCOS and controls irrespective of obesity (Figure 1A, Table 2). On the contrary, citramalic acid, glycine, gluconic acid lactone, phenylalanine and palmitoleic acid were different when comparing non-obese with obese women

irrespective of PCOS (Figure 1B, Table 2) and glycerol, lactic acid and oleic acid were influenced by obesity, but this influence was different in the PCOS and control groups (Figure 2, Table 2).

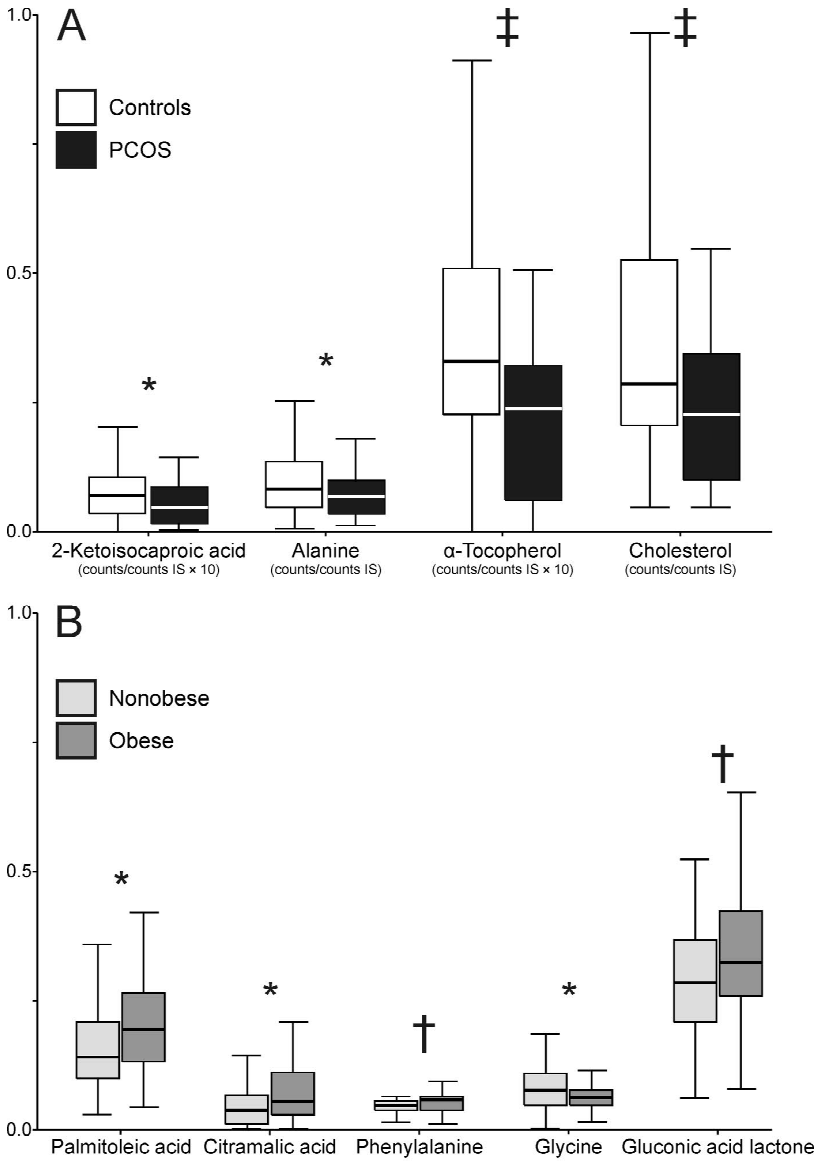


Fig. 1. Metabolites showing differences in the m/zRT intensities normalized to internal standard succinic-d4 acid. The box plot includes the median (horizontal line) and the interquartile range, and the whiskers indicate the minimum and maximum data values, unless outliers are present, in which case the whiskers extend to a maximum of 1.5 times the interquartile range. (A), Differences between patients with polycystic ovarysyndrome (PCOS) and controls, irrespective of obesity. (B), Differences between nonobese and obese women, irrespective of PCOS status. *P _ 0.05; †P _ 0.01; ‡P _ 0.005.

We found statistically significant interactions between PCOS and obesity in the plasma levels of 3 metabolites also affected by obesity – lactic acid, glycerol and oleic acid –and of 3 metabolites – adipic acid, linoleic acid and glyceric acid – that were affected by the interaction between obesity and PCOS but showed no overall significant effects for any of these independent variables (Figure 2, Table 2). Of the 6 metabolites showing statistically significant interactions among PCOS and obesity, glycerol and adipic, linoleic and glyceric acid were reduced in non-obese PCOS patients and increased in obese PCOS patients when compared to their non-obese and obese counterparts (Figure 2). Lactic acid was increased in non-obese subjects compared with obese women considering PCOS patients and controls as a whole, yet this difference was caused by an increase in non-obese PCOS patients (Figure 2). Conversely, oleic acid was increased in obese subjects compared with non-obese women considering PCOS patients and controls as a whole, but oleic acid concentrations decreased in non-obese PCOS patients and increased in obese PCOS patients when compared to non-obese and obese controls, respectively (Figure 2).

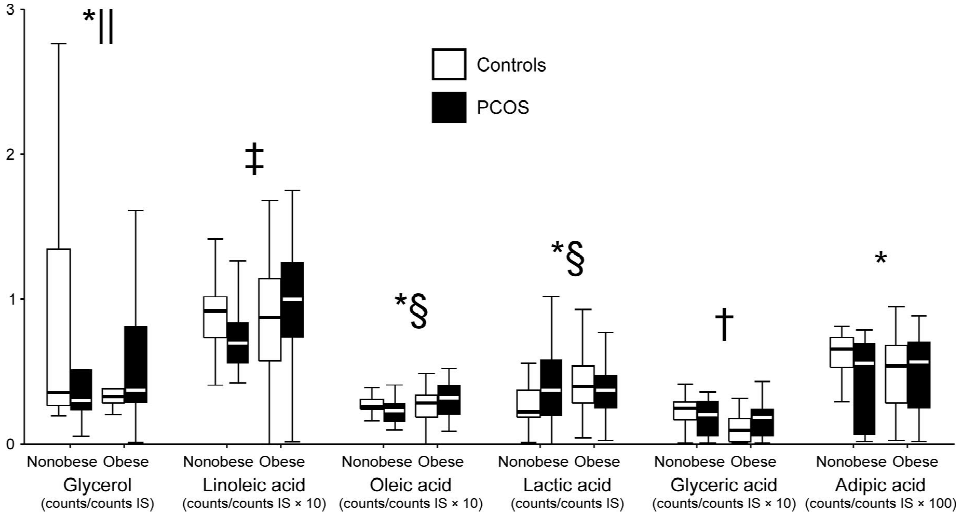


Fig. 2. Metabolites showing statistically significant interactions between PCOS and obesity, meaning that the effects of PCOS were different in obese and nonobese women, or that the effects of obesity were not the same in the PCOS and control groups. Data are m/zRT intensities normalized to internal standard succinic-d4 acid. The box plot includes the median (horizontal line) and the interquartile range, and the whiskers indicate the minimum and maximum data values, unless outliers are present, in which case the whiskers extend to a maximum of 1.5 times the interquartile range. *P <0.05, †P < 0.01, and ‡P _ 0.005 for the interaction between PCOS and obesity; §P < 0.05 and ||P < 0.001 for the effect of obesity irrespective of PCOS status.

Metabolite	Fold change	Quantitative ion (m/z)	RT (min)	Fiehn/NIST match (% Probability NIST or Net Fiehn)
Patients with PCOS vs controls				
2-Ketoisocaproic acid	↑1.5	216	9.15	90
Alanine[b]	↑1.3	188	11.2	97
α-Tocopherol[b]	↑1.6	502	27.45	88
Cholesterol	↑1.5	458	27.56	97
Obese vs non-obese women				
Palmitoleic acid[b]	↓1.2	311	18.72	87
Phenylalanine[b]	↓1.2	120	13.58	77
Glycine[b]	↑1.2	248	10.48	77
Citramalic acid[b]	↓1.3	247	12.72	71
Gluconic acid lactone	↓1.2	217	17.0	69
Lactic acid	↓1.0	191	7.01	99
Glycerol[b]	↑1.4	205	10.05	91
Oleic acid[b]	↓1.4	339	20.53	99
Interaction between PCOS and obesity				
Lactic acid	-	191	7.01	99
Glyceric acid	-	292	10.85	95
Adipic acid[b]	-	275	13.05	72
Glycerol[b]	-	205	10.05	91
Linoleic acid[b]	-	337	20.43	96
Oleic acid[b]	-	339	20.53	99

Table 2. Results of untargeted GC-MS approach to the study of the interplay between obesity and PCOS in premenopausal women. a Metabolites showing statistically significant differences after GC-MS raw data, normalized to internal standard succinic-d4 acid, and rank transformed. Metabolites were submitted to 2-way ANOVA introducing PCOS status and obesity as independent variables (see also Figs. 1 and 2).b Identification was unequivocally confirmed by comparison of retention times and spectral data to the corresponding pure standard compounds. Fold changes are also presented (1indicates increased relative areas of the quantitative ion with respect to its corresponding control group;2indicates decreased relative areas). We used AMDIS and both NIST and Fiehn libraries to identify the most likely metabolites from the features differentially present as a function of PCOS status, obesity, or their interaction.

DISCUSSION

Our present results provide novel data indicating substantial metabolic heterogeneity in PCOS, modulated mostly by its association with obesity.

The increase in plasma long-chain fatty acids such as linoleic and oleic acid and of glycerol found in the obese patients with PCOS, as well as the increase in oleic and palmitoleic acid found in obese women irrespective of PCOS, suggests increased lipolysis possibly secondary to impaired insulin action at adipose tissue [100,101]. The increase in circulating free fatty acids may contribute further to the insulin resistance associated with both obesity and PCOS [78], and to the association of PCOS and obesity with the metabolic syndrome and non-alcoholic fatty liver disease [102].

These findings are in conceptual agreement with the metabolic signature previously associated with obesity [103] that also includes increased circulating branched-chain aminoacids (BCAA) [103]. Peripheral insulin resistance leads to an increase in circulating BCAA concentrations because their utilization in tissues such as muscle requires conserved insulin signaling. In a situation of insulin resistance, BCAA are used for gluconeogenesis through pyruvate transamination into alanine, thereby contributing to glucose intolerance [103,104]. The decreased glycine levels in our obese patients may also indicate increased utilization for gluconeogenesis. Furthermore, an increase in BCAA catabolism may explain the increase in phenylalanine concentrations in our obese women, because large neutral amino acids

such as phenylalanine compete with BCAA for transport into mammalian cells [105].

Yet the changes observed in the metabolic profile of our women with PCOS, especially in the non-obese sample, strongly suggest that insulin resistance is not universal in these patients and that the increase in serum insulin levels may result from entirely different mechanisms. First, the changes in glycerol and long-chain fatty acids in non-obese women with PCOS were just the opposite than those found in the obese PCOS patients, as these metabolites were decreased in non-obese patients with PCOS compared with the non-obese controls, suggesting suppression of lipolysis. This finding requires conserved insulin sensitivity in adipose tissue, and possibly the presence of increased insulin concentrations, although an impairment in catecholamine-induced lipolysis in subcutaneous tissue is also characteristic on non-obese women with PCOS [100,106,107] and might have contributed to the reduced lipolysis suggested by our present results. Second, the decrease in 2-ketoisocaproic acid levels in patients with PCOS suggests decreased transamination of leucine in the first step of BCAA catabolism – which is common to leucine, iso-leucine and valine – instead of the increased transamination previously suggested as a metabolic footprint of obesity and diabetes [103,104]. Under normal conditions, alanine arising from BCAA nitrogen accounts for 25% of gluconeogenesis from amino acids [108]. These data, together with the decreased alanine levels found in patients with PCOS, suggests that BCAA are being used for protein synthesis in these women, and not for gluconeogenesis as happens in obesity and diabetes [103,104]. However, the actual meaning of change

in alanine levels is ambiguous because a decrease in its plasma concentration has been also associated to increased gluconeogenesis in diabetic db-/db- mice [109].

Although a certain degree of insulin resistance was present in our patients with PCOS because these women remained normoglycemic in the presence of hyperinsulinemia, these results suggest that peripheral insulin resistance did not dominate the picture in a significant number of PCOS patients, especially in the non-obese. Suppression of lipolysis in adipose tissue and protein synthesis in tissues such as muscle require conserved insulin signaling [110]. The finding of increased plasma lactic acid levels in our non-obese patients with PCOS further suggests insulin-stimulated glucose uptake and consumption in the muscle of these patients. In conceptual agreement, oocytes from non-obese patients with PCOS show increased glucose and pyruvate consumption during overnight *in vitro* maturation, a characteristic that is not present in oocytes when patients had been treated with the insulin-lowering drug metformin [111]. These results may suggest that there is no resistance to the actions of insulin in PCOS oocytes, and that their increased glucose consumption may result from exposure to increased insulin levels.

We may also speculate that the putative increase in BCAA usage with respect to that of non-hyperandrogenic women may be related to the anabolic effects of androgens on protein synthesis, for example in muscle and bone mass [112], coupled to the increased circulating insulin levels that characterize both lean and obese women with PCOS [84]. Of note, a large body of evidence supports the possibility of insulin action

varying between metabolic responses and between target tissues in women with PCOS [78,113], in conformity with our present results.

For this explanation to be plausible, the increased insulin levels of lean women with PCOS must not result only from peripheral insulin resistance. There are scientifically sound data supporting this possibility/ on the one hand, euglycemic hyperinsulinemic clamp studies demonstrated that in non-obese women with PCOS a reduced hepatic clearance of insulin, and not only peripheral insulin resistance, is a major contributor to their increased circulating insulin concentrations whereas obese PCOS patients presented with peripheral insulin resistance [114]; on the other hand, myotubes from patients with PCOS may show normal insulin sensitivity [113]. Therefore, an increase in insulin levels coupled to the lack of resistance for some insulin actions or lack of insulin resistance in some non-peripheral target tissues, may explain an increase in utilization of BCAA and glucose in muscle and perhaps in other tissues, as suggested by the metabolic profile of our non-obese patients with PCOS.

The decreased concentrations of the lipid-soluble vitamin α-tocopherol could contribute to the increased oxidative stress previously associated with PCOS [115]. On the contrary, the finding of decreased free cholesterol concentrations in our patients with PCOS is relatively unexpected as we were not able corroborate any change in plasma lipid levels by classical analytical procedures in these women. Furthermore, we have no clear explanation for the changes in other metabolites such as citramalic, adipic acid, gluconic acid lactone or glyceric acid, although the last two metabolites (which are products of glucose and glycerol

oxidation, respectively) may represent artifacts resulting from the high temperatures reached during gas chromatography.

In summary, our present results indicate substantial metabolic heterogeneity in PCOS, suggesting that insulin resistance and hyperinsulinemia contribute to the pathogenesis of this disorder, but that the later is not necessarily secondary to the former. Non-obese women with PCOS present a metabolic footprint compatible with a predominant effect of hyperinsulinemia in the absence of peripheral insulin resistance at adipose tissue and muscle. This might be related to the previously described decrease in hepatic clearance of insulin in such women [114] and to the heterogeneity in insulin action between metabolic responses and between target tissues characteristic of women with PCOS [78,113]. On the contrary, obese women with PCOS present with a metabolic profile in which the consequences of insulin resistance, possibly related to abdominal adiposity [84], predominate.

However, the most plausible scenario is that patients with PCOS may present throughout a whole spectrum of relative contributions of both pathogenetic mechanisms to hyperinsulinemia, with obesity playing a definite role as inductor of insulin resistance. The possibility that hyperinsulinemia may occur in the absence of universal insulin resistance in some women with PCOS, especially when abdominal adiposity and / or obesity are not marked, should be considered when designing diagnostic and therapeutic strategies for the management of this prevalent disorder.

REFERENCES

1. Asuncion M, Calvo RM, San Millan JL, Sancho J, Avila S, Escobar-Morreale HF. A prospective study of the prevalence of the polycystic ovary syndrome in unselected Caucasian women from Spain. J Clin Endocrinol Metab 2000;85/2434-8.

2. Escobar-Morreale HF, San Millan JL. Abdominal adiposity and the polycystic ovary syndrome. Trends Endocrinol Metab 2007;18/266-72.

3. Azziz R, Carmina E, Dewailly D, Diamanti-Kandarakis E, Escobar-Morreale HF, Futterweit W, et al. The Androgen Excess and PCOS Society criteria for the polycystic ovary syndrome/ the complete task force report. Fertil Steril 2009;91/456-88.

4. Ehrmann D, Barnes R, Rosenfield R, Cavaghan M, Imperial J. Prevalence of impaired glucose tolerance and diabetes in women with polycystic ovary syndrome. Diabetes Care 1999;22/141-6.

5. Legro RS, Kunselman AR, Dodson WC, Dunaif A. Prevalence and predictors of risk for type 2 diabetes mellitus and impaired glucose tolerance in polycystic ovary syndrome/ a prospective, controlled study in 254 affected women. J Clin Endocrinol Metab 1999;84/165-9.

6. Luque-Ramirez M, Alpanes M, Escobar-Morreale HF. The determinants of insulin sensitivity, beta-cell function, and glucose tolerance are different in patients with polycystic ovary syndrome than in women who do not have hyperandrogenism. Fertil Steril 2010;94/2214-21.

7. Salley KE, Wickham EP, Cheang KI, Essah PA, Karjane NW, Nestler JE. Glucose intolerance in polycystic ovary syndrome--a position

statement of the Androgen Excess Society. J Clin Endocrinol Metab 2007;92/4546-56.

8. Wild RA, Carmina E, Diamanti-Kandarakis E, Dokras A, Escobar-Morreale HF, Futterweit W, et al. Assessment of cardiovascular risk and prevention of cardiovascular disease in women with the polycystic ovary syndrome/ a consensus statement by the Androgen Excess and Polycystic Ovary Syndrome (AE-PCOS) Society. J Clin Endocrinol Metab 2010;95/2038-49.

9. Escobar-Morreale HF, Insenser M, Corton M, San Millán JL, Peral B. Proteomics and genomics/ a hypothesis-free approach to the study of the role of visceral adiposity in the pathogenesis of the polycystic ovary syndrome. Proteomics Clin Appl 2008;2/444-56.

10. Corton M, Botella-Carretero JI, Benguria A, Villuendas G, Zaballos A, San Millan JL, et al. Differential gene expression profile in omental adipose tissue in women with polycystic ovary syndrome. J Clin Endocrinol Metab 2007;92/328-37.

11. Corton M, Botella-Carretero JI, Lopez JA, Camafeita E, San Millan JL, Escobar-Morreale HF, Peral B. Proteomic analysis of human omental adipose tissue in the polycystic ovary syndrome using two-dimensional difference gel electrophoresis and mass spectrometry. Hum Reprod 2008;23/651-61.

12. Insenser M, Martinez-Garcia MA, Montes R, San-Millan JL, Escobar-Morreale HF. Proteomic analysis of plasma in the polycystic ovary syndrome identifies novel markers involved in iron metabolism, acute-phase response, and inflammation. J Clin Endocrinol Metab 2010;95/3863-70.

13. Dettmer K, Aronov PA, Hammock BD. Mass spectrometry-based metabolomics. Mass Spectrom Rev 2007;26/51-78.

14. Bain JR, Stevens RD, Wenner BR, Ilkayeva O, Muoio DM, Newgard CB. Metabolomics applied to diabetes research: moving from information to knowledge. Diabetes 2009;58:2429-43.

15. Escobar-Morreale HF, Sanchon R, San Millan JL. A prospective study of the prevalence of nonclassical congenital adrenal hyperplasia among women presenting with hyperandrogenic symptoms and signs. J Clin Endocrinol Metab 2008;93:527-33.

16. Zawadzki JK, Dunaif A. Diagnostic criteria for polycystic ovary syndrome: Towards a rational approach. In: Dunaif A, Givens JR, Haseltine FP, Merriam GR, eds. Polycystic ovary syndrome, Vol. 4. Boston: Blackwell Scientific Publications, 1992:377-84.

17. The Rotterdam ESHRE/ASRM-sponsored PCOS consensus workshop group. Revised 2003 consensus on diagnostic criteria and long-term health risks related to polycystic ovary syndrome (PCOS). Hum Reprod 2004;19:41-7.

18. Azziz R, Carmina E, Dewailly D, Diamanti-Kandarakis E, Escobar-Morreale HF, Futterweit W, et al. Position statement: criteria for defining polycystic ovary syndrome as a predominantly hyperandrogenic syndrome: an Androgen Excess Society guideline. J Clin Endocrinol Metab 2006;91:4237-45.

19. Martinez-Garcia MA, Luque-Ramirez M, San-Millan JL, Escobar-Morreale HF. Body iron stores and glucose intolerance in premenopausal women: role of hyperandrogenism, insulin resistance and genomic

variants related to inflammation, oxidative stress and iron metabolism. Diabetes Care 2009;32:1525-30.

20. Matsuda M, DeFronzo RA. Insulin sensitivity indices obtained from oral glucose tolerance testing: comparison with the euglycemic insulin clamp. Diabetes Care 1999;22:1462-70.

21. Palazoglu M, Fiehn O. Metabolite identification in blood plasma using GC-MS and the Agilent Fiehn GC-MS metabolomics RTL library. Agilent Application Note 2009.

22. Dunn WB, Broadhurst D, Begley P, Zelena E, Francis-McIntyre S, Anderson N, et al. Procedures for large-scale metabolic profiling of serum and plasma using gas chromatography and liquid chromatography coupled to mass spectrometry. Nat Protoc 2011;6:1060-83.

23. Ek I, Arner P, Bergqvist A, Carlstrom K, Wahrenberg H. Impaired adipocyte lipolysis in nonobese women with the polycystic ovary syndrome: a possible link to insulin resistance? J Clin Endocrinol Metab 1997;82:1147-53.

24. Mozaffarian D, Cao H, King IB, Lemaitre RN, Song X, Siscovick DS, Hotamisligil GS. Circulating palmitoleic acid and risk of metabolic abnormalities and new-onset diabetes. Am J Clin Nutr 2010;92:1350-8.

25. Dunaif A. Insulin resistance and the polycystic ovary syndrome: mechanism and implications for pathogenesis. Endocr Rev 1997;18:774-800.

26. Setji TL, Holland ND, Sanders LL, Pereira KC, Diehl AM, Brown AJ. Nonalcoholic steatohepatitis and nonalcoholic Fatty liver disease in young women with polycystic ovary syndrome. J Clin Endocrinol Metab 2006;91:1741-7.

27. Newgard CB, An J, Bain JR, Muehlbauer MJ, Stevens RD, Lien LF, et al. A branched-chain amino acid-related metabolic signature that differentiates obese and lean humans and contributes to insulin resistance. Cell Metab 2009;9:311-26.

28. Suhre K, Meisinger C, Doring A, Altmaier E, Belcredi P, Gieger C, et al. Metabolic footprint of diabetes: a multiplatform metabolomics study in an epidemiological setting. PLoS ONE 2010;5:e13953.

29. Fernstrom JD. Branched-chain amino acids and brain function. J Nutr 2005;135:1539S-46S.

30. Ek I, Arner P, Ryden M, Holm C, Thorne A, Hoffstedt J, Wahrenberg H. A unique defect in the regulation of visceral fat cell lipolysis in the polycystic ovary syndrome as an early link to insulin resistance. Diabetes 2002;51:484-92.

31. Faulds G, Ryden M, Ek I, Wahrenberg H, Arner P. Mechanisms behind lipolytic catecholamine resistance of subcutaneous fat cells in the polycystic ovarian syndrome. J Clin Endocrinol Metab 2003;88:2269-73.

32. Layman DK. The role of leucine in weight loss diets and glucose homeostasis. J Nutr 2003;133:261S-7S.

33. Altmaier E, Ramsay SL, Graber A, Mewes HW, Weinberger KM, Suhre K. Bioinformatics analysis of targeted metabolomics--uncovering old and new tales of diabetic mice under medication. Endocrinology 2008;149:3478-89.

34. Biolo G, Declan Fleming RY, Wolfe RR. Physiologic hyperinsulinemia stimulates protein synthesis and enhances transport of selected amino acids in human skeletal muscle. J Clin Invest 1995;95:811-9.

35. Harris SE, Maruthini D, Tang T, Balen AH, Picton HM. Metabolism and karyotype analysis of oocytes from patients with polycystic ovary syndrome. Hum Reprod 2010;25:2305-15.

36. Notelovitz M. Androgen effects on bone and muscle. Fertil Steril 2002;77 Suppl 4:S34-41.

37. Dunaif A, Segal KR, Shelley DR, Green G, Dobrjansky A, Licholai T. Evidence for distinctive and intrinsic defects in insulin action in polycystic ovary syndrome. Diabetes 1992;41:1257-66.

38. Ciaraldi TP, Aroda V, Mudaliar S, Chang RJ, Henry RR. Polycystic ovary syndrome is associated with tissue-specific differences in insulin resistance. J Clin Endocrinol Metab 2009;94:157-63.

39. Ciampelli M, Fulghesu AM, Cucinelli F, Pavone V, Caruso A, Mancuso S, Lanzone A. Heterogeneity in beta cell activity, hepatic insulin clearance and peripheral insulin sensitivity in women with polycystic ovary syndrome. Hum Reprod 1997;12:1897-901.

40. Fenkci V, Fenkci S, Yilmazer M, Serteser M. Decreased total antioxidant status and increased oxidative stress in women with polycystic ovary syndrome may contribute to the risk of cardiovascular disease. Fertil Steril 2003;80:123-7.

2.2.1.2 Metabolomics untargeted approach reveals impaired transport reverse of cholesterol in young lean PCOS patients

INTRODUCTION

Polycystic ovary syndrome (PCOS) is the most common endocrine disorder in women of reproductive age. The diagnosis of PCOS is based on three main criteria: (i) hyperandrogenism, (ii) oligo-amenorrhea, and (iii) the observation of polycystic ovaries on a sonogram [1]. While the mecanisms leading to the development of PCOS are still not completely understood, PCOS has been associated with multifactorial causes including the interaction of genetic variants, environmental factors and life-style [2]. In addition, increasing evidences indicate that insulin resistance and compensatory hyperinsulinaemia may play an important role in the pathophysiology of PCOS [3]. In fact, long-term studies have shown that the risk of glucose intolerance is approximately 5 to 10-fold higher in PCOS relative to healthy women, and this risk does not appear to be limited to a single ethnic group [4,5]. Consequently, PCOS women constitute a high-risk group for developing type 2 diabetes [3,6].

Metabolomics enables the characterization of small molecules that are direct signatures of biochemical activity. The implementation of untargeted metabolomic approches, therefore, may facilitate the identification of novel molecules implicated in the pathophysiology of PCOS including the development of common metabolic disorders such as type 2 diabetes.

The aim of this study is to apply our metabolomic approach to discover new biomarkers of PCOS. We analyzed serum samples of young

PCOS patients and healthy controls using LC-QTOF MS and NMR to investigate metabolic alterations characteristics of PCOS.

MATERIALS AND METHODS

Study population

The study girls were recruited among patients consecutively seen in the Adolescent Endocrinology Unit of Sant Joan University Hospital, Barcelona, Spain. The study population consisted of 12 girls with hyperinsulinemic androgen excess and 14 control girls.

Inclusion criteria for PCOS girls were: 1) hyperinsulinemia, defined as fast-inginsulinemia above 15 U/ml and/or a peak insulinemia above 150 U/ml, and/or mean insulinemia above 84 U/ml on a 2-h oral glucose tolerance test (18); and 2) the presence of both clinical and biochemical androgen excess, as defined by the following: hirsutism score above 8 (Ferriman-Gallwey), amenorrhea (no menses for 3 months) or oligomenorrhea (menstrual cycles 45 d); and high circulating levels of androstenedione or testosterone in the follicular phase (d 3–7) or after 2 months of amenorrhea.

Exclusion criteria were: evidence of anemia, thyroid dysfunction, bleeding disorder, Cushing syndrome, or hyperprolactinemia; glucose intolerance; diabetes mellitus; late-onset adrenal hyperplasia; abnormal electrolytes; abnormal screening of liver or kidney function; use of medication affecting gonadal or adrenal function, or carbohydrate or lipid metabolism. Pregnancy risk was a particular exclusion criterion that was not only taken into account at study start, but was also maintained throughout the study in the PioFluMet subgroup.

Untargeted metabolomic experiment

Untargeted metabolomics analysis on serum samples was performed using two analytical platforms: NMR and LC-ESI-QTOF; each serum sample was split into two aliquots and run in parallel using the two analytical platforms. For the NMR measurement 250 µL of serum were mixed with 250 µL of phosphate buffer (0.75 mM Na2HPO4 adjusted at pH 7.4, and 20% D2O to provide the field frequency lock). The final solution was transferred to a 5 mm NMR tube and kept refrigerated at 4ºC in the autosampler until the analysis. 1H-NMR spectra were recorded at 310 K on a Bruker Avance III 600 spectrometer operating at a proton frequency of 600.20 MHz using a 5 mm CPTCI triple resonance (^{1}H, ^{13}C, ^{31}P). Three different ^{1}H-NMR pulse experiments were performed for each sample: 1) Nuclear Overhauser Effect Spectroscopy (NOESY)-presaturation sequence to suppress the residual water peak; 2) Carr-Purcell-Meiboom-Gill sequence (CPMG, spin-spin T2 relaxation filter) with a total time filter of 410 ms to attenuate the signals of serum macro-molecules to a residual level; 20 ppm spectral width and a total of 64 transients collected into 64 k data points, and 3) Diffusion-edited pulse sequence with bipolar gradients along with longitudinal eddy-current delay (LED) to further estimate the serum lipoprotein profile according to our described methodology [7].

The second aliquot was used for LC-MS analysis. 30 µL of serum sample was mixed with 120 µL of cold ACN:H2O with 1% MPA and 0.1% formic acid (previously filtered). Samples were vortexed vigorously for 30 seconds and stored at –20ºC for 2 hours to enable protein precipitation. Subsequently, samples were centrifuged 15 minutes at 4ºC and 15000

rpm and the supernetant was transferred to a LC-MS vial. Samples were injected in a UHPLC system (1290 Agilent) coupled to a quadrupole time of flight (QTOF) mass spectrometer (6550 Agilent Technologies) operated in positive electrospray ionization (ESI+) mode. Metabolites were separated using either C18-RP (ACQUITY UPLC HSS T3 1.8 µL, Waters) and HILIC (ACQUITY UPLC BEH 1.7 µL, Waters) chromatography at a flow rate of 0.4 mL/min. The solvent system in C18-RP was A= 0.1% formic in water, and B= 0.1% formic in acetonitrile. The solvent system in HILIC was A= 50mM NH4OcA in water, and B= ACN. The injection volume was 2 µL. MS/MS data of the metabolites of interest was collected using the same equipment and runing the samples again. Finally, those metabolites confirmed by MS/MS using QTOF quantified again using a triple quadrupole (QQQ) MS (6490 Agilent Technologies). Quality control samples (QC) consistig of pooled serum samples of all patients entering the study were used. QC samples were injected before the first study samples and then, were analyzed periodically after five-study samples. Furthermore, real samples were randomize to reduce systematic error associated with instrumental drift.

Proteomic experiment

Protein precipitation was carried out adding 10% of pure trichloroacetic acid (TCA) to 10 µL of serum. Samples were vortexed vigorously and incubated on ice for 1 hour. Then, samples were centrifuged at 4ºC and 14000 rpm for 15 minutes and the supernatant was discarded. 800 µL of cold acetone (-20ºC) were added to the pellet and proteins were suspended and incubated overnight (-20ºC). Samples were centrifuged and the supernatant discarded again. This step was

repeated and the pellet was air-dried. Pellet was resuspended with 100 µL of urea (7M), thiourea (2M) and CHAPS (4%) buffer. 10 µL of this solution were added to 30 µL of Laemmli Buffer (4x). 40 µL of such solution was applied to a home-made on 12% acrylamide/bis-acrylamide SDS-PAGE gels.

Proteins were Coomassie stained. Bands of interest were manually excised from 1D SDS-PAGE gels and were destained and washed with 25 mMAmBicfor 15 min followed by a wash with acetonitrile for 15 min. These washes were twice repeated and samples were finally dehydrated with 100% acetonitrile and dried in a speed Vack. For the cysteine carbamidomethylation, dried gel was placed at 56 ºC for 1 h in a reducing solution containing 10 mM DTT and 50mM ammonium bicarbonate. Alkylation of the cysteines was achieved by incubation of the gel for 30 min in the dark with 55 mMiodoacetamide in 25 mM ammonium bicarbonate buffer. Gel pieces were alternately washed with 25 mMAmBic and 25 mMAmBic with acetonitrile, and finally dehydrated with 100% acetonitrile and dried under vacuum. All gel pieces were incubated with 12.5 ng/µl sequencing grade trypsin (Roche Molecular Biochemicals) in 25 mMAmBic overnight at 37°C. After digestion, the supernatants were separated. Peptides were further extracted from the gel pieces into 50% ACN, 0.1% trifluoroacetic acid. For each extraction, samples were incubated for 10 min in and orbital shaker. All extracts were pooled, and the volume was reduced by SpeedVac. In order to obtain a suitable sample for mass spectrometry analysis the pellet was resuspendedin 25µL of 0.1% TFA in water, and was desalted and

concentrated using C18 ZipTips (Millipore). Tryptic peptides were sequentially eluted with 5 µL of 70% acetonitrile with 0.1% TFA in water.

MALDI-TOF MS

Samples were spotted following the dried-droplet method where 1 µl reconstituted in-gel digest sample was spotted initially on a BigAnchorChip target plate (BrukerDaltonics), followed by 1 µl of matrix (10 mg/ml -cyano-4-hydroxycinnamic acid matrix (BrukerDaltonics) in 50% ACN, 0.1% trifluoroacetic). Sample and matrix mixture was left to dry at room temperature. Mass spectra were obtained on aUltrafleXtreme (BrukerDaltonics, Bremen, Germany) matrix-assisted laser desorption ionization–tandem time of flight (MALDI-TOF/TOF) mass spectrometer. Mass spectra were recorded in the positive ion reflectronmodein the mass range of 700–3500 Da.Operating conditions were as follows: ion source 1 = 25.00 kV, ion source 2 = 24.40 kV, lens voltage = 8.50 kV, reflector voltage = 26.45 kV, optimized pulsed ion extraction time = 130ns, matrix suppression = 500 Da. All mass spectra were externally calibrated using a standard peptide mixture (Bruker); calibration was considered good when a value below 1 ppm was obtained. For MS analysis, 1500 single-shot spectra were accumulated by recording 50-shot_spectra at 10 random positions using fixed laser attenuation.

Database Search

For peptide mass fingerprinting analysis, ProteinScape software (Bruker) supported by the Mascot search engine (Matrix Science) was used with the following parameters: SWISS-PROT non-redundant database filtered by homo sapiens taxonomy,two missed cleavage

permission, 50-ppm measurement tolerance. Carbamidomethylation of cysteines was set as a fixed modification and methionine oxidation was set as a variable modification.Positive identifications were accepted with a Mascot score higher than that corresponding to a *P* value of 0.05.

Ethics

This clinical study was registered as ISRCTN12871246 and conducted in Sant Joan de Déu University Hospital (Barcelona, Spain), without support from industry, after approval by the Institutional Review Board of Sant Joan de Déu University Hospital, and after written informed consent by each patient.

Data analysis and statistical methods

The acquired CPMG NMR spectra were phased, baseline corrected and referenced to the chemical shift of the α-glucose anomeric proton doublet at 5.23 ppm. Pure compound references in BBioref AMIX (Bruker), HMDB and Chenomx databases were used for metabolite identification. After baseline correction, intensities of each 1H-NMR regions identified in the CPMG 1D-NMR spectra were integrated for each sample entering the study using the AMIX 3.8 software package (Bruker, GmBH).

LC-MS (C18 and HILIC ESI+ mode) data were processed using the XCMS software [24] (version 1.36.0) to detect and align features. A feature is defined as a molecular entity with a unique m/z and a specific retention time. XCMS analysis of these data provided a matrix containing the retention time, m/z value, and integrated peak area of each feature for every serum sample extraction discussed above. The tab-separated text files containing LC-MS data were imported into Matlab where QC

samples were used to filter analytical variation (section 2.1.2). Then univariate statistical analysis was performed using robust statistics (Yuen test). Differentially regulated metabolites (fold>1.5) that passed our statistical criteria (p-value<0.01) were characterized by LC-qTOF MS/MS and identified using Metlin database. Finally, identified metabolites by matching MS/MS spectra with databases were additionally quantified by LC-QQQ MS in MRM mode.

RESULTS

Biochemical changes

Different biochemical parameters involved in PCOS disease were monitored in PCOS and control women. Table 1 shows mean values and p-values for each condition. PCOS women presented statistically increased levels of testosterone, dehydroepiandrosterone sulfate (DHEAS) and leptin. Levels of superoxide dismutase (SOD) were statistically decreased in PCOS patients.

	CTR	PCOS	p-value
Age	17.2 ± 0.4	16.3 ± 0.4	0.15
BW	3336.4 ± 48.7	3154.2 ± 142.4	0.46
GA	39.5 ± 0.3	39.9 ± 0.4	0.48
BW SDS	0.3 ± 0.1	-0.2 ± 0.4	0.58
Wt	58.8 ± 1.8	58.2 ± 1.2	0.92
Ht	163.9 ± 1.3	160.0 ± 1.5	0.06
BMI	21.8 ± 0.6	22.8 ± 0.5	0.23
BMI SDS	0.2 ± 0.2	0.5 ± 0.2	0.16
WBC	7.3 ± 0.3	7.8 ± 0.5	0.26
Neutrophils	4.1 ± 0.3	4.4 ± 0.5	0.63
Lymphocytes	2.2 ± 0.1	2.5 ± 0.22	0.41
N/L	1.9 ± 0.2	2.0 ± 0.4	0.51
oGTT	89.1 ± 1.5	85.4 ± 2.0	0.14
Testosterone	32 ± 2.4	64.2 ± 10.2	0.05
DHEAS	222.1 ± 27.8	280.8 ± 31.5	0.03
Leptin	13.9 ± 2.3	20.9 ± 2.7	0.05
PCRUS	0.7 ± 0.2	1.1 ± 0.2	0.14
SOD	6.1 ± 0.3	5.4 ± 0.2	0.03

Table 1. Biochemical parameters. Data are represented as mean±standard error of the mean (SEM). P-values are calculated from a robust Yuen test. BW: birth weight, GA: gestational age, BW SDS: Birth weight standard desviation, Wt: weight, Ht: height, BMI: body mass index, BMI SDS_ body mass index standard desviation, WBC: white blood cells, N/L: Neutrophils/Lymphocytes, AST: aspartate transaminase, ALT: alanine aminotransferase, oGTT: oral glucose tolerance test, DHEAS: dehydroepiandrosterone sulphate PCRUS: protein c reactive ultrasensible, SOD: superoxid dismutase

Lipoprotein Profile

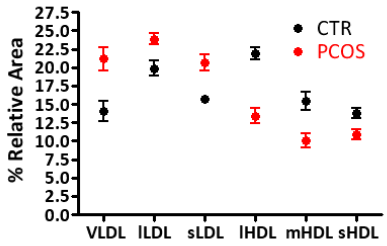
Our metabolomic study includes a lipoprotein profile obtained based on an existing NMR procedure previously described [7]. In brief, 1H NMR spectra were recorded at 310 K on a Bruker Avance III 600 spectrometer operating at a proton frequency of 600.20 MHz (14.1 T). We used the double stimulated echo (DSTE) pulse program with bipolar gradient

pulses and a longitudinal eddy current delay (LED). The relaxation delay was 2 s, the free induction decays (FIDs) were collected into 64K complex data points and 32 scans were acquired on each sample. The gradient pulse strength was increased from 5 to 95% of the maximum strength of 53.5 Gauss cm-1 in 32 steps. The squared gradient pulse strength was linearly distributed.

The methyl signal was surface fitted using eight Lorentzian functions as previously reported [7]. The average diffusion coefficients obtained for each function were used to calculate an average size by means of the Stokes-Einstein equation. According to these NMR-derived lipoprotein sizes, functions 2 to 8 were associated with 1 VLDL, 2 LDL, and 4 HDL lipoprotein subclasses, respectively. For simplification, functions 7 to 8 were grouped to obtain the small HDL subclass. Thus, our NMR-derived lipoprotein subclasses were defined as VLDL, large LDL, small LDL, large HDL, medium HDL, and small HDL..

Figure 2 shows the serum lipoprotein profile of PCOS and control women. Levels of VLDL, small LDL and large LDL are increased in PCOS patients relative to controls. In contrast, the abundance of large, medium and small HDL subfractions are decreased in PCOS. Figure 2B shows a descriptive table with the percentage of variation for each subfraction. Every lipoprotein subfraction showed significant differences between PCOS and control women; however, the greatest variation is associated with large HDLs. All together, our NMR characterization of lipoprotein subclasses in serum reveals a dyslipidemic profile in women with PCOS.

A)



B)

	p-value	% variation
VLDL	1.29E-2	50
Large LDL	2.64E-2	20
Small LDL	2.45E-2	31
Large HDL	1.55E-5	-63
Medium HDL	5.80E-3	-52
Small HDL	2.33E-2	-20

Figure 1. Lipoprotein profile measured by NMR. (A) Row-wise normalized areas showed as mean±sem. VLDL: very low-density lipoprotein, lLDL: large low-density lipoprotein, sLDL: small low-density lipoprotein, lHDL: large high-density lipoprotein, mHDL: medium high-density lipoprotein, sHDL: small high-density lipoprotein (B) p-values obtained after robust Yuen test and percentage of variation for each subfraction.

Metabolic changes induced by PCOS

Untargeted metabolomics using NMR and LC-qTOF MS reveals different disregulated metabolites in women with PCOS. We quantified the relative abundance of metabolites in serum by comparing the normalized integrated area of NMR resonances and the integrated area of each feature using the XCMS software [9] for mass spectrometry data. In addition, we confirmed the quantitative and statistical differences of every metabolite by means of their quantitation using LC-QqQ MS in MRM mode. Table 2 shows p-values and percentage of variation of those metabolites that vary significantly between healthy and women with PCOS.

	p-value	% variation
Methionine	5.04E-05	-344
Methionine Sulfoxide	1.77E-03	40
5-oxoproline	4.98E-03	39
Taurine	1.53E-02	28
Glu-Cys	1.56E-04	37
Glu-Glu	2.70E-02	35
Glutamate	4.04E-03	33
GSH	3.60E-02	34
GSH/GSSG	2.30E-02	29
Glu-Taurine*	3.03E-07	73
Glycine	3.10E-02	-18

Table 2. Metabolites quantified by LC-QqQ MS in MRM mode. P-values (robust Yuen test) and percentages of variation are shown. Glu: glutamate, cys: cysteine, GSH: glutathione, GSH/GSSG: glutathione/oxidated glutathione. A negative % variation value indicates down-regulation in PCOS relative to control women, and vice versa. *Glu-taurine could not be quantified by LC-QqQ in MRM mode due to the lack of pure standard. Values for Glu-taurine are from the LC-qTOF profiling.

Most metabolites identified in our untargeted metabolomic approach are involved in the γ-glutamyl cycle (Figure 1), which is a metabolic cycle for transporting amino acids into cells and plays a role for the synthesis and degradation of glutathione. Dipeptides (Glu-Glu, Glu-Tau or Glu-Cys), amino acids (glutamate, 5-oxoproline or taurine), GSH and the GSH/GSSG ratio were significantly increased in women with PCOS. In contrast, glycine levels were found significantly decreased in PCOS. All together, these results point to increasing activation of γ-glutamyl cycle in PCOS, which could be attributed to the necessity of compensating oxidative stress in these women.

Moreover, our results show an inverse behaviour of methionine and methionine sulfoxide. Oxidation of methione residues in apo-AI have been associated with impaired reverse of cholesterol in HDL [10], which

would partly explain the differences in HDL subclasses by NMR. Since apo-AI is the major protein component of HDLs and one of the most abundant proteins in normal human plasma [11], we hypothesize that women with PCOS present greater oxidation of methionine residues in apo-AI relative to healthy women, and that the increased levels of free methionine sulfoxide in PCOS serum may result from the turnover and degradation of these apo-AI proteins.

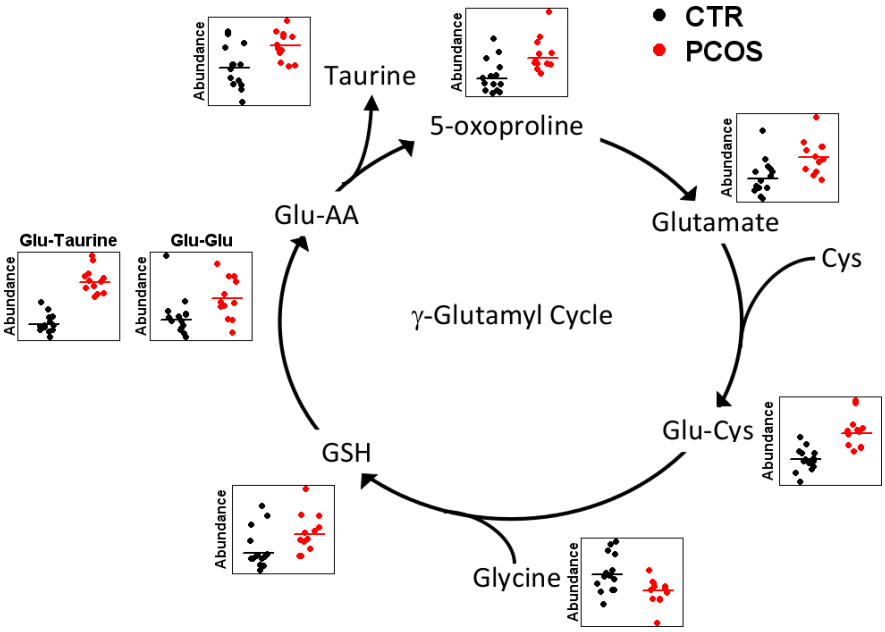


Figure 2. Identified metabolites from the γ-glutamyl cycle. The plots show abundance (i.e., intensity) of metabolite in each sample and trimmed mean (control in black and PCOS in red).

Quantitative analysis of the MetOx/Met ratio in apo AI by MALDI-TOF MS

To test the hypothesis that women with PCOS suffer from increased oxidation of methionine residues in apo-AI, we performed a proteomic

study measuring the ratio of methionine sulfoxide/methionine in apo-AI. Concretely, we were interested in the ratio of Met-148 residue, because Met-148 residue is involved in the LCAT activation domain of apo-AI [12].

Our proteomic analysis by MALDI-TOF MS showed a significant increase of the methionine oxidation/methionine ratio in women with PCOS relative to controls (robust Yuen test, p=6.3x10-4). Furthermore, a positive and statistically significant correlation between methionine sulfoxide/methionine in apo-AI and free methionine sulfoxide in serum was observed (Figure 3).

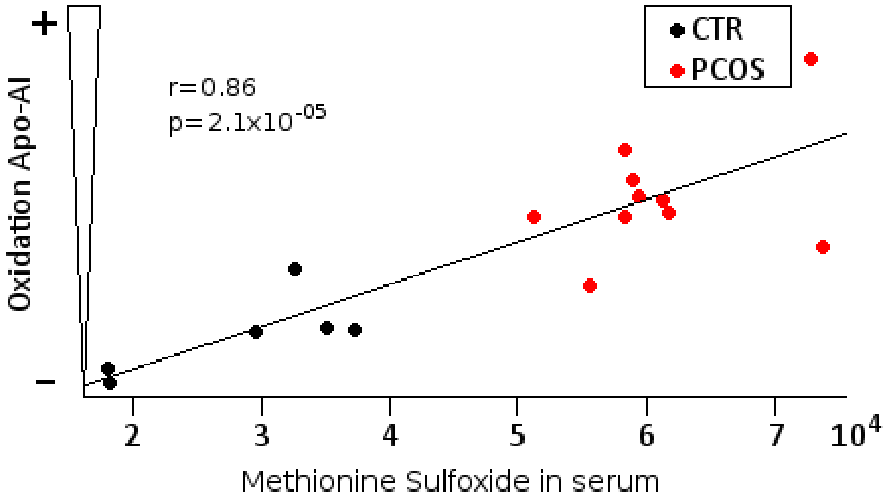


Figure 3. Correlation between methionine sulfoxide in serum and the oxidation state of apo-AI. Oxidation of apo-AI is calculated as the ratio of oxidized Met-148/Met-148 obtained by MALDI-TOF MS.

These results suggest that levels of methionine sulfoxide in serum are directly associated with the oxidation level of apo-AI. Moreover, we found a negative correlation between the number of large HDL particles and methionine sulfoxide in serum (Figure 4), which reinforced our

hypothesis that levels of methionine sulfoxide in serum reflect HDL oxidation .

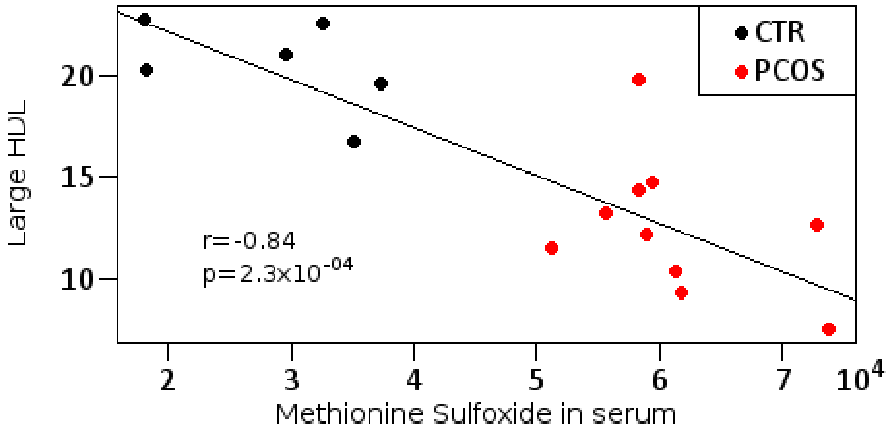


Figure 4. Negative correlation between methionine sulfoxide in serum and number of large HDL particles.

These results together suggest that levels of methionine sulfoxide in serum, MetOx-148/Met-148 ratio in apo-AI and levels of large HDL may represent a novel axis to study dyslipidemia.

DISCUSSION

Oxidative stress is present in many disorders related to insulin resistance including PCOS. It is the result of an imbalance between the production of reactive oxigen species (ROS) and their inactivation by antioxidant systems in cells and tissues [13]. Low levels of antioxidant enzymes such as SOD in women with PCOS may contribute to the detrimental effects of ROS. Our results indicate that women with PCOS activate the biosynthesis of glutathione (GSH) through the γ-glutamyl cycle to compensate their oxidative imbalance (i.e., redox status) [14].

This is reflected in greater levels of GSH (and GSH/GSSG ratio) in PCOS. An imbalance in the antioxidant protective mechanism has being identified as a common factor in atherosclerotic diseases [15], diabetes [16], metabolic syndrome [17] and PCOS [18].

This oxidative state can cause oxidative modifications in lipoproteins [19]. It is well known that PCOS women commonly presents dyslipidemia which usually includes low HDL cholesterol, high LDL cholesterol and high triglyceride concentrations [20]. Women with PCOS show an increased cardiovascular risk due in part to this dyslipidemia [20]. However, the cause of an atherogenic profile in PCOS women is still unknown [21]. Oxidative modifications are considered an initial step in lipoprotein conversion into more atherogenic particles [10]. Increased oxidation of plasma LDL and accumulation of oxidatively modified LDL (oxLDL) in macrophages in the arterial wall are characteristic of the early stages of atherogenesis [22]. In contrast, HDL cholesterol is generally associated with lower risk of cardiovascular disease [23]. Concretely, HDL-C is generally associated with atheroprotective properties, which include mediation of reverse cholesterol trasnsport. The functional status of HDL is closely linked to its primary protein component, apo-AI, an abundant apolipoprotein whose plasma concentrations are inversely correlated with the incidence of coronary artery disease [24].

However, recent studies have cast some doubt on the “good cholesterol” HDL hypothesis [10,25,26]. Shao et. al. demonstrated that oxidation of methionine residues in apo-AI impairs reverse cholesterol transport by HDL [10]. Specifically, the oxidation of Met-148 in apo-AI impairs apo-AI’s ability to activate lecithin cholesterol acyltransferase

(LCAT) [10]. LCAT is the enzyme responsible for transforming nascent HDL into spherical HDL particles containing a central hydrophobic core of cholesteryl esters and an outer layer composed of apo-AI and phospholipids. Therefore, oxidation of Met-148 in apo-AI impairs reverse cholesterol transport [10].

Our NMR lipoprotein study shows lower number of large (i.e., mature) HDL particles in PCOS relative to control women. Together with the increased oxidation of Met-148 residues in apo-AI in women with PCOS, our results suggest reduced ability to activate LCAT in PCOS, which would prevent the formation of mature HDL particles. To the best of our knowledge, loss of LCAT activity has not been reported in women with PCOS. Rajkhowa, et al. [20] measured the concentration of apo-AI in HDL particles, but they did not find any statistical difference between control and PCOS women. We did not measure the total amount of apo-AI, however, our (and previous) results would indicate that HDL functionality is not regulated by protein concentration but by post-translational modifications such as oxidation in methionine residues. Therefore, the atherogenic profile of women with PCOS would result from an elevated oxidative stress that, in turn, impact HDL functionality through oxidation of apo-AI. Moreover, increased oxidation of methionine residues in apo-AI has been described previously in type 1 diabetes [27], but no relationship has been done with LCAT activity in those patients.

The same dyslipidemia associated with PCOS is also characteristic of metabolic syndrome, which is associated with type 2 diabetes and cardiovascular disease. Women with PCOS have a higher risk of

developing type 2 diabetes [6] and cardiovascular disease [28]. We should emphasize, however, that our patients do not present any evidence of glucose intolerance (p-value>0.05) at the moment of the study. So we can consider them as pre-diabetic. Therefore, the measurement of methionine sulfoxide and HDL subfractions together with the ratio of methionine oxidation in apo-AI can be a novel early biomarker to predict the development of metabolic syndrome and its complications. Such measurements can complement other diagnostic tools to identify patients at a risk for developing diabetes.

CONCLUSIONS

Women with PCOS typically present cardiovascular risk factors such as obesity, hyperinsulinemia, impaired glucose tolerance, hyperandrogenemia or dyslipidemia that can lead to serious endocrinological, metabolic and cardiovascular disorders. However, women diagnosed with PCOS at an early stage, such as those of our study, do not present all these adverse features. Therefore, elevated levels of methionine sulfoxide in serum, oxidized methionine-148 in apo-AI, and the lipoprotein profile might constitute earlier biomarkers of metabolic syndrome.

REFERENCES

1. Witchel SF, Recabarren SE, Gonzalez F, Diamanti-Kandarakis E, Cheang KI, et al. (2012) Emerging concepts about prenatal genesis, aberrant metabolism and treatment paradigms in polycystic ovary syndrome. Endocrine 42: 526-534.

2. Insenser M, Montes-Nieto R, Murri M, Escobar-Morreale HF (2013) Proteomic and metabolomic approaches to the study of polycystic ovary syndrome. Molecular and Cellular Endocrinology 370: 65-77.

3. Seli E, Duleba AJ (2002) Should patients with polycystic ovarian syndrome be treated with metformin? Proven and potential benefits. Human Reproduction 17: 2230-2236.

4. Pelusi B, Gambineri A, Pasquali R (2004) Type 2 diabetes and the polycystic ovary syndrome. Mierva Ginecology 56: 41-51.

5. García-Romero G, Escobar-Morreale HF (2006) Hyperandrogenism, Insulin Resistance and Hyperinsulinemia as Cardiovascular Risk Factors in Diabetes Mellitus. Current Diabetes Reviews 2: 39-49.

6. Nestler JE (2002) Should patients with polycystic ovarian syndrome be treated with metformin? An enthusiastic endorsement. Human Reproduction 17: 1950-1953.

7. Mallol R, Rodriguez MA, Heras M, Vinaixa M, Canellas N, et al. (2011) Surface fitting of 2D diffusion-edited H-1 NMR spectroscopy data for the characterisation of human plasma lipoproteins. Metabolomics 7: 572-582.

8. Vinaixa M, Samino S, Saez I, Duran J, Guinovart J, et al. (2012) A Guideline to Univariate Statistical Analysis for LC/MS-Based Untargeted Metabolomics-Derived Data. Metabolites 2: 775-795.

9. Smith CA, Want EJ, O'Maille G, Abagyan R, Siuzdak G (2006) XCMS: Processing mass spectrometry data for metabolite profiling using Nonlinear peak alignment, matching, and identification. Analytical Chemistry 78: 779-787.

10. Shao B, Cavigiolio G, Brot N, Oda M, Heinecke J (2008) Methionine oxidation impairs reverse cholesterol transport by apoprotein A-1. PNAS 105: 12224-12229.

11. Fisher E, Feig J, Hewing B, Hazen S, Smith J (2012) High-Density Lipoprotein Function, Dysfunction, and Reverse Cholesterol Transport. Arteriosclerosis, Thrombosis, and Vascular Biology 32: 2813-2820.

12. Sorci-Thomas M, Thomas M (2002) The effects of altered apolipoprotein A-I structure on plasma HDL concentration. Trends in Cardiovascular Medicine 12: 121-128.

13. Sabuncu T, Vural H, Harmna M, Harma M (2001) Oxidative stress in polycystic ovary syndrome and its contribution to the risk of cardiovascular disease. Clinical Biochemistry 34: 407-413.

14. Zhang H, Forman H, Choi J (2005) γ-Glutamyl Transpeptidase in Glutathione Biosynthesis. Methods in Enzimology 401: 468-483.

15. Stocker R, Keaney J (2005) New insights on oxidative stress in the artery wall. Journal Thromb Haemost 3: 1825-1834.

16. Martitim A, Sanders R, Watkins J (2003) Diabetes, oxidative stress, and antioxidants: a review. Journal of Biochemical Molecular Toxicology 17: 24-38.

17. Roberts C, Sindhu K (2009) Oxidative stress and metabolic syndrome. Life Sci 84: 705-712.

18. Sundaram R, Bhaskar A, Vijayalingam S, Viswanathan M, Moahn R, et al. (1996) Antioxidant status and lipid peroxidation in type II diabetes mellitus with and without complications. Clin Sci 90: 255-260.

19. Kaysen G, Eiserich J (2004) The role of oxidative stress-altered lipoprotein structure and function and microinflammation on

cardiovascular risk in patients with minor renal dysfunction. Journal American Society Nephrology 15: 538-548.

20. Rajkhowa M, Neary R, Kumpatla P, Game F, Jones P, et al. (1997) Altered Composition of High Density Lipoproteins in Women with the Polycystic Ovary Syndrome. Journal of Clinical Endocrinology and Metabolism 82: 3389-3394.

21. Dokras A (2013) Cardiovascular disease risk in women with PCOS. Steroids 78: 773-776.

22. Jialal I, Devaraj S (1996) The role of oxidized low density lipoprotein in atherogenesis. Journal of Nutrition 126: 1053S-1057S.

23. Valkenburg O, Steegers-Theunissen R, Smedts H, Dallinga-Thie G, Fauser B, et al. (2008) A more atherogenic serum lipoprotein profile is present in women with polycystic ovary syndrome: a case-control study. Journal Clinical Endocrinology Metabolism 93: 470-476.

24. Borja MS, Zhao L, Hammerson B, Tang C, Yang R, et al. (2013) HDL-apoA-I Exchange: Rapid Detection and Association with Atherosclerosis. PLoS ONE 8: e71541.

25. Voight B, Peloso G, Orho-Melander M (2012) lasma HDL cholesterol and risk of myocardial infarction: a mendelian randomisation study. The Lancet 380: 572-580.

26. The AIM-HIGH investigator (2011) Niacin in Patients with Low HDL Cholesterol Levels Receiving Intensive Statin Therapy. New England Journal of Medicine 365: 2255-2267.

27. Brock J, Jenkins A, Lyons T, LKlein R, Yim E, et al. (2008) Increased methionine sulfoxide content of apoA-I in type 1 diabetes. Journal of Lipid Research 49: 847-855.

28. Dahlgren E, Janson P (1994) Polycystic ovary syndrome: long-term metabolic consequences. International Journal of Gynaecology Obstet 44: 3-8.

2.2.2 Metabolic responses to pharmacological therapies in PCOS

Metabolomics reveals reduction of metabolic oxidation in women with polycystic ovary syndrome after pioglitazone-flutamide-metformin polytherapy

SUMMARY

Polycystic ovary syndrome (PCOS) is a variable disorder characterized by a broad spectrum of anomalies, including hyperandrogenemia, insulin resistance, dyslipidemia, body adiposity, low-grade inflammation and increased cardiovascular disease risks. Recently, a new polytherapy consisting of low-dose flutamide, metformin and pioglitazone in combination with an estro-progestagen resulted in the regulation of endocrine clinical markers in young and non-obese PCOS women. However, the metabolic processes involved in this phenotypic amelioration remain unidentified. In this work, we used NMR and MS-based untargeted metabolomics to study serum samples of young non-obese PCOS women prior to and at the end of a 30 months polytherapy receiving low-dose flutamide, metformin and pioglitazone in combination with an estro-progestagen. Our results reveal that the treatment decreased the levels of oxidized LDL particles in serum, as well as downstream metabolic oxidation products of LDL particles such as 9- and 13-HODE, azelaic acid and glutaric acid. In contrast, the radiuses of small dense LDL and large

HDL particles were substantially increased after the treatment. Clinical and endocrine-metabolic markers were also monitored, showing that the level of HDL cholesterol was increased after the treatment, whereas the level of androgens and the carotid intima-media thickness were reduced. Significantly, the abundance of azelaic acid and the carotid intima-media thickness resulted in a high degree of correlation. Altogether, our results reveal that this new polytherapy markedly reverts the oxidant status of untreated PCOS women, and potentially improves the pro-atherosclerosis condition in these patients.

INTRODUCTION

Hyperandrogenemia, insulin resistance, a state of low-grade inflammation, body adiposity and a pro-atherogenic lipid profile are usually present in adolescents and young women with polycystic ovary syndrome (PCOS) [1,2,3]. Previous reports have evidenced that the PCOS phenotype can also concur with primary alterations of lipid metabolism involving increased levels of oxidized LDL (oxLDL) particles [4,5,6]. Accordingly, PCOS may accelerate the development of a cardiovascular-risk profile even in the absence of clinical signs of atherosclerosis. Whether this adverse pro-atherogenic profile and the enhanced oxidant status may increase the risk for cardiovascular disease remains unclear [7,8,9].

Most current pharmacological therapies are addressed to ameliorate menstrual irregularities and cosmetic issues. There is a

clear need, however, for treatments that also improve endocrine-metabolic markers associated with this disorder. Low-dose flutamide (Flu, a pure androgen receptor blocker) and metformin (Met, an insulin sensitizer) in combination with an estro-progestagen, is a polytherapy that reduces total and abdominal fat, decreases the lean mass deficit, and attenuates the abnormal pattern of adipokines in young and non-obese PCOS women [1]. Recently, it has been reported that the addition of low-dose pioglitazone (Pio, a PPARγ agonist) to the aforementioned polytherapy confers further reductions of visceral fat and carotid intima-media thickness (IMT), and increases further circulating high molecular-weight (HMW) adiponectin [10,11,12,13]. Although the phenotypic evidences demonstrate endocrine-metabolic improvements in a wide spectrum of long-term health markers, the molecular mechanisms underlying such polytherapy remain to be elucidated.

An approach to explore the metabolic changes in PCOS women caused by this new treatment is metabolomics, defined as the metabolic complement of functional genomics. Metabolomics enables the characterization of endogenous small molecules that serve as direct signatures of biochemical activity and therefore are easier to correlate with phenotype. With the ultimate goal of the comprehensive metabolome coverage, there is an overriding need for analytical methodologies able to produce comprehensive metabolite profiles from complex biological samples. However, due to the huge physico-chemical diversity of metabolites, there is not

a unique analytical technology able to cope with the whole metabolome. Mass spectrometry (both GC-MS and LC-MS) and NMR have demonstrated to be complementary analytical technologies in metabolomics-based studies [14,15,16], allowing to expand the number of metabolites that can be comprehensively covered in an untargeted experiment. Advantages and disadvantages of each analytical technique and their current status in the field of metabolomics are excellently reviewed elsewhere [17,18,19,20].

Here we present a longitudinal study to reveal the mechanism of action of the low-dose Pio/Flu/Met polytherapy using metabolomics. We analyzed using LC-MS, GC-MS and NMR-based metabolomics serum samples of twelve young, non-obese women diagnosed with PCOS, prior to and at the end of a 30 months treatment with low-dose Pio/Flu/Met in combination with an oral estro-progestagen. Our results show that at the end of the treatment there are marked changes in the size of different lipoprotein particles, in conjunction with downstream metabolic oxidation products of LDL particles.

METHODS

Participants

The study population consisted of twelve young, non-obese women (age, 19.6 ± 0.4 yr; BMI, 22.3 ± 0.9 Kg/m2) diagnosed with PCOS, participating in a randomized study [11], and whose metabolomic profiles were assessed at baseline and after 30

months of treatment. Over 30 months, all women received the same therapy for 24/28 days/ pioglitazone (7.5 mg/d) at breakfast, and metformin (850 mg/d), flutamide (62.5 mg/d) and a contraceptive (ethinylestradiol 20 µg/d plus drospirenone 3 mg/d; Yasminelle, Schering) at dinner time.

The inclusion criteria were: 1) hyperinsulinemia on a standard 2-h oral glucose tolerance test, defined as peak insulin levels >150 U/mL and/or mean serum insulin >84µU/mL; 2) ovarian hyperandrogenism, as defined by each of the following symptoms: hirsutism (Ferriman & Gallwey score >8); amenorrhea (no menses for >3 months) or oligomenorrhea (duration of cycles > 45 days); biochemical androgen excess, as judged by circulating androstenedione, total testosterone or free androgen index [FAI, testosterone x 100/sex hormone-binding globulin (SHBG)]; 17-OH-progesterone hyperresponse (>160 ng/dL) to GnRH agonist stimulation (leuprolide acetate 500 µg subcutaneously) [10].

The main exclusion criteria were: BMI <17 Kg/m2 or >29 Kg/m2; evidence of thyroid dysfunction; Cushing syndrome or hyperprolactinemia; glucose intolerance; personal history of diabetes mellitus; late-onset adrenal hyperplasia; abnormal liver or kidney function; abnormal blood counts or serum electrolytes; and treatment with an oral contraceptive or another medication known to affect gonadal or adrenal function, carbohydrate or lipid metabolism.

Patients were selected based on available serum samples pre- and post-treatment. PCOS patients receiving placebo for 30

months did not meet ethical requirements and were not included in the clinical study. Blood sampling was performed at baseline and after 30 months on a cyclic off-treatment day (4/28 days). For metabolomics analysis, serum samples (600 µL) were obtained allowing plasma to clot at room temperature for 30 min. After centrifugation at 4 °C at 10,000 g for 10 min, samples were maintained at - 80 °C until further analysis.

Experimental Procedures

Clinical and endocrine-metabolic variables, carotid IMT, body composition [by dual-energy X-ray absorptiometry (DXA)] and abdominal fat partitioning [by magnetic resonance imaging (MRI)] were assessed prior to and at the end of the treatment as previously described [9,10,11,12]. Sampling was performed in the follicular phase of the cycle, or after 2 months of amenorrhea. Hirsutism was graded according to the Ferriman and Gallwey score [11]. Fasting blood glucose, serum insulin, LDL- and HDL-cholesterol, sex hormone-binding globulin (SHBG), testosterone, androstenedione and dehydroepiandrosterone-sulfate (DHEAS), carotid IMT, body composition and abdominal fat partitioning were measured as previously described [10,11,12,13].

Untargeted metabolomics analysis on serum samples was performed using three different analytical platforms: NMR, GC-MS and LC-ESI-MS TOF; each serum sample was split into three aliquots and run in parallel using the three analytical platforms.

For the NMR measurement 300 µL of serum were mixed with 300 µL of phosphate buffer (0.75 mM Na_2HPO_4 adjusted at pH 7.4,

and 20% D_2O to provide the field frequency lock). The final solution was transferred to a 5 mm NMR tube and kept refrigerated at 4ºC in the autosampler until the analysis. ^{1}H-NMR spectra were recorded at 310 K on a Bruker Avance III 600 spectrometer® operating at a proton frequency of 600.20 MHz using a 5 mm CPTCI triple resonance (^{1}H, ^{13}C, ^{31}P). Three different ^{1}H-NMR pulse experiments were performed for each sample/ 1) Nuclear Overhauser Effect Spectroscopy (NOESY)-presaturation sequence to suppress the residual water peak; 2) Carr-Purcell-Meiboom-Gill sequence (CPMG, spin-spin T2 relaxation filter) with a total time filter of 410 ms to attenuate the signals of serum macro-molecules to a residual level; 20 ppm spectral width and a total of 64 transients collected into 64 k data points, and 3) Diffusion-edited pulse sequence with bipolar gradients along with longitudinal eddy-current delay (LED) to further estimate the serum lipoprotein profile according to our recent described methodology [21].

The second aliquot was used for GC-MS analysis according to Agilent's specifications [22]. 100 µL of serum were spiked with 20 µL of internal standard solution (1 µg/µL succinic-d4 acid; Sigma-Aldrich). After protein precipitation using 900 µL of cold methanol/water (8/1 v/v) samples were centrifuged 10 minutes at 4ºC, and 200 µL of the supernatant were spiked with 20 µL of myristic acid-d27 (Sigma Aldrich) used for retention time lock. Samples were then lyophilized, and dissolved and incubated in 50 µL of methoxyamine in pyridine (0.3 µg/µL) during 16 hours at room temperature. Derivatization by silylation reagents was done

using 30 µL of N-methyl-N-trimethylsilyltrifluoroacetamide with 1% trimethylchlorosilane (MSTFA + 1% TMCS, Sigma) during 1 hour at room temperature. Samples were automatically injected into a GC–MS system (HP 6890 Series gas chromatograph coupled to a mass selective detector model 5973) equipped with a J&W Scientific DB 5-MS+DG stationary phase column (30 m × 0.25 mm i.d., 0.1 µm film) (Agilent Technologies). The injector temperature was set at 250 ºC, and the helium carrier flow rate was kept constant at 1.1 mL/min. The column temperature was held at 60 ºC for 1 min, then increased to 325ºC at a rate of 10 ºC/min and held at 325ºC for 10 min. The detector operated in the electron impact ionization mode (70 eV) and mass spectra were recorded after a solvent delay of 4 min with 2.46 scans per second (mass scanning range of m/z 50–600; threshold abundance value of 50 counts). The source temperature and quadrupole temperature were 230 and 150 °C, respectively.

The third aliquot was filtered through a 0.22 µm nylon membrane filter and directly injected in a HPLC system (1200 series, Agilent Technologies) coupled to a time-of-flight (TOF) mass spectometer (6210 Agilent Technologies) operated either in positive (ESI+) or negative (ESI-) electrospray ionization in full scan mode. Serum extractions were separated using a Kinetex C18, 2.6 µm, 150 x 2.1 mm, 100 A (Phenomenex, Torrance, CA) at a flow rate of 0.4 mL/min. The solvent system was: A= 0,1% formic acid in water; B= 0,1% formic acid in acetonitrile (ACN). The gradient profile started linearly from 2% to 20% buffer B in 3 min and was

followed by another linear gradient from 20% to 100% buffer B in 18 min and hold for 7 min at 100% buffer B.The injection volume was 15 µL. The instrument was set to acquire over the m/z range 80-1000 with an acquisition rate of 1.3 spectra/second. MS/MS data of the metabolites of interest was collected using an HPLC-ESI QqQ system (6410, Agilent Technologies) using identical LC conditions.

Quality control samples (QCs) consisting of pooled serum samples of all patients entering the study were used. In our LC-MS platform, QCs were injected before the first study samples were analyzed and then periodically after 5-study samples. For GC-MS, QCs were injected periodically after 10-study samples. In addition, to begin with the chromatographic analysis, injection of 3 blank runs were performed both in LC-MS and GC-MS. Furthermore, samples entering the study were entirely randomizedto reduce systematic error associated with instrumental drift.

Ethics

This clinical study was registered as ISRCTN12871246 and conducted in Sant Joan de Déu University Hospital (Barcelona, Spain), without support from industry, after approval by the Institutional Review Board of Sant Joan de Déu University Hospital, and after written informed consent by each patient.

Data analysis and statistical methods

The acquired CPMG NMR spectra were phased, baseline-corrected and referenced to the chemical shift of the α-glucose anomeric proton doublet at 5.22 ppm. Pure compound references

in BBioref AMIX (Bruker); HMDB and Chenomx databases were used for metabolite identification. After baseline correction, intensities of each ^{1}H-NMR regions identified in the CPMG 1D-NMR spectra were integrated for each sample entering the study using the AMIX 3.8 software package (Bruker, GmBH). To identify discriminating markers, intensities of each of the identified spectral regions in the untreated PCOS serum spectra were compared against the same spectral regions on their treated counterparts using principal component analysis (PCA) of the auto-scaled within-subject variation dataset derived from multilevel simultaneous component analysis (MSCA) [23] and the non-parametric Wilcoxon-rank summed test. LC-MS (ESI+ and ESI- mode) and GC-MS data were processed using the XCMS software [24](version 1.6.1) to detect and align features. A feature is defined as a molecular entity with a unique m/z and a specific retention time. XCMS analysis of these data provided a matrix containing the retention time, *m/z* value, and integrated peak area of each feature for every serum sample extraction discussed above. The tab-separated text files containing GC-MS data were imported into Matlab where normalization to internal standard succinic acid-d4 was also performed. QCs were always projected in a PCA model together with the study samples to verify that technical issues do not mask biological information. Basal PCOS samples and their treated counterparts were compared using the integrated peak area of each feature via PCA of the auto-scaled within-subject variation dataset derived from MSCA and the non-parametric

Wilcoxon-rank summed test, and assigning a fold value to indicate the level of differential regulation due to the 30-months treatment. Differentially regulated metabolites (fold>2) that were statistically significant ($p<0,05$) detected by LC-MS were characterized by MS/MS using a LC-QqQ instrument. Differentially regulated metabolites detected by GC-MS were identified using the NIST and Fiehn mass spectral libraries. In addition, the retention time of pure standards were confirmed. Data (pre-) processing, data analysis, and statistical calculations were performed in Matlab (Matlab version 6.5.1, Release 13). The MSCA matlab code was downloaded and adapted from www.bdagroup.nl.

RESULTS

Low-dose Pio/Flut/Met polytherapy improves biochemical long-term health markers in PCOS patients

A series of biochemical parameters were initially monitored in PCOS patients prior to and at the end of the 30 months treatment. The results summarized in Table 1 indicate that the treatment caused a broad spectrum of biochemical adjustments, including a marked reduction in serum concentrations of androgens such as testosterone (-41±9%, $p=0.0026$) and androstenedione (-35±5%; $p=0.003$), whereas insignificant changes in body weight were measured. The carotid IMT was markedly reduced (-33±3%; $p=3.0\times10^{-5}$) after the treatment, and it was accompanied by a significant augment of HDL cholesterol levels in serum (34±5%, $p=0.005$). In addition, some PCOS patients decreased their visceral

fat mass considerably after the treatment. Of note, markers of liver dysfunction such as transaminases and lactate dehydrogenase remained unaltered (data not shown). Overall, these changes are clinically associated with a phenotypic improvement in PCOS patients after the treatment, and complement the metabolomic analysis.

	At baseline	At 30 mo	p
N	12	12	
Age (yr)	19.6 ± 0.4		-
BMI (Kg/m2)	22.3 ± 0.9	22.6 ± 1.0	0.9770
Score F&G	16.1 ± 1.2	8.1 ± 0.6	0.0001
Total cholesterol (mg/dl)	164 ± 7	184 ± 8	0.0606
HDL-cholesterol (mg/dl)	50 ± 2	68 ± 4	0.0055
LDL-cholesterol (mg/dl)	101 ± 6	96 ± 5.0	0.7726
Triglycerides (mg/dl)	65 ± 5	99 ± 9	0.0093
Testosterone (ng/dl)	76 ± 6	47 ± 7	0.0026
SHBG (µg/dL)	1.0 ± 0.1	4.6 ± 0.2	$4.0X10^{-5}$
FAI	8.8 ± 1.6	1.0 ± 0.2	$4.0X10^{-5}$
Androstendione (ng/dl)	440 ± 44	279 ± 23	0.0030
DHEAS (µg/dl)	264 ± 34	207 ± 31	0.2365
Carotid IMT (mm)	0.46 ± 0.02	0.31 ± 0.01	$3.5X10^{-5}$
Subcutaneus fat mass (cm2)	151.8 ± 19.5	155.6 ± 21.6	0.8852
Visceral fat mass (cm2)	50.3 ± 4.5	40.4 ± 6.9	0.0783
Visceral- to-subcutaneous ratio	0.37 ± 0.04	0.28 ± 0.04	0.1484

Table 1. Endocrine-metabolic markers, carotid IMT and abdominal fat partitioning at baseline and after 30 months of low dose Pio/Flu/Met polytherapy.Values are mean 6 SEM; BMI, body mass index; SHBG, sex hormone-binding globulin; DHEAS, dehydroepiandrosterone-sulfate; FAI, free androgen index; IMT, Intima-Media Thickness.

Multivariate Data Analysis of PCOS Serum Samples

Given the longitudinal nature of our study, we used a multilevel simultaneous component analysis (MSCA) method [23] to examine

independently different types of variation in the NMR and MS-based metabolomic data, namely, variation between patients and variation in time within patients as a result of the polytherapy. MSCA enables to split the variation within-subjects that accounts for the variation before and after the treatment for each patient, from the variation between-subjects that accounts for the biological variation (i.e., genotype differences) in two different data matrices. The variation within-subjects is represented in the PCA plot of Figure 1. The PC1/PC2 scores plot of the ^{1}H-NMR CPMG spin-echo experiment shown in Figure 1A reveals a clear clustering along PC1 (~57% of the variance), which accounts for the post-treatment variation of each patient. Figure 1B depicts the bar loading plot for PC1, the higher the absolute bar value the higher the influence of such variable in the variation induced by the treatment. Hence, PCOS serum samples at the end of the treatment were characterized by rather larger values of choline-containing molecules and diminished values of 1,2-propanediol and lysine, among others. Similar PC1/PC2 scores plot was observed with the GC-MS and LC-MS derived data (Figure 1C and Figure 1F), confirming a profound metabolic variation within patients due to the 30 months polytherapy. Loadings bar plot were also studied and some of the features implicated in this variation were depicted as boxplots in Figures 1D-H. Altogether our MSCA model set apart the biological variation between patients and highlighted within-patient variation due to the Pio/Flu/Met polytherapy, showing clear differences in the relative

concentration of specific compounds in the serum of PCOS patients. Accordingly, we focused in the characterization of the most discriminating compounds of the MSCA model.

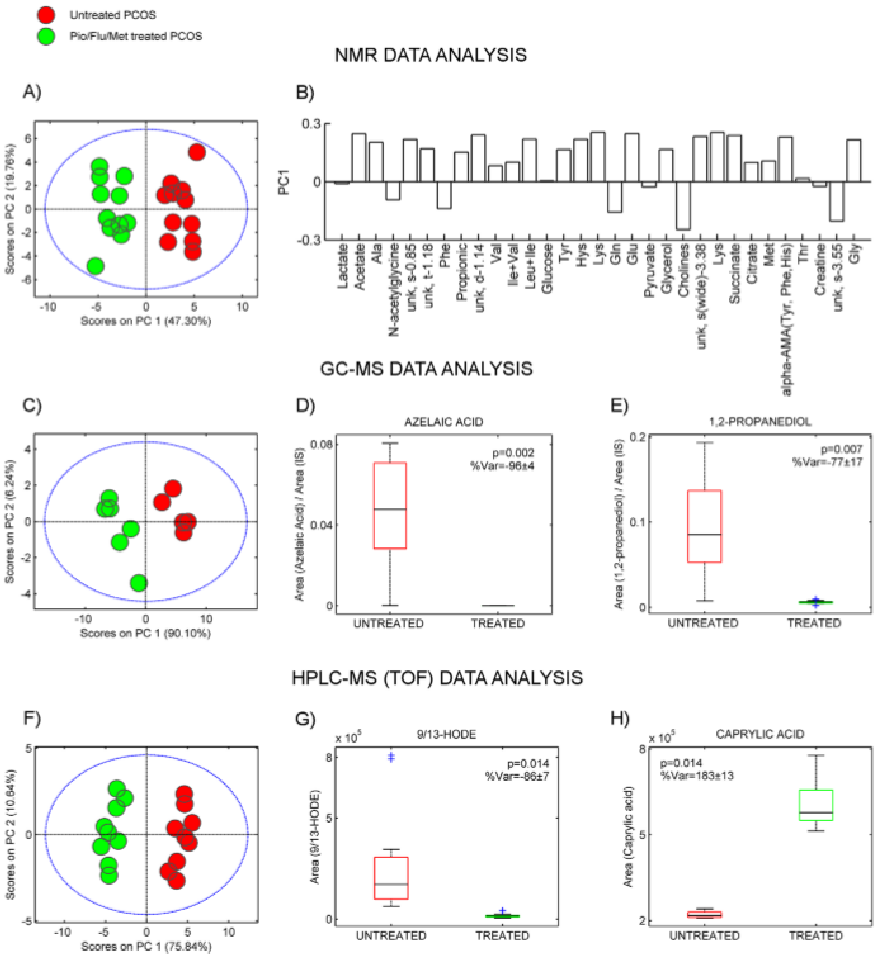


Figure 1. Multivariate modelling of 1N-NMR, GC-MS and LC-ESI-TOF MS data. (A) PC1/PC2 scatter scores plot and (B) PC1 loading bar plot of PCA calculated on the within-subject matrix derived from the MSCA modelling of the 32 selected spectral regions identified in the 1H-NMR CPMG serum spectra of untreated (red markers) and treated (green markers) PCOS patients. (C) PC1/PC2 scatter scores plot of PCA calculated on the withinsubject matrix derived from the MSCA modelling of GC-MS data. Boxplots of (D) azelaic acid and (E) 1,2-proanediol, the two metabolites corresponding to the most discriminating features along the corresponding PC1 loadings bar plot. (F) PC1/PC2 scatter scores plot of PCA

calculated on the within-subject matrix derived from the MSCA modelling of LC-ESI-TOF MS data. Boxplots of (G) 9- and 13-HODE, and (H) caprylic acid, the two metabolites corresponding to the most discriminating features along the PC1 loadings bar plot. P-values derived from Wilcoxon rank-summed pairmatched comparison of untreated and treated PCOS patients. Mean 6 sem of the percentage of variation are also indicated.

Analysis of serum ^{1}H-NMR spectra

The CPMG ^{1}H NMR serum spectra are composed of overlapped resonances from low molecular weight metabolites such as amino acids or lactate and T_2 lipoprotein attenuated signals. Figure 2A depicts representative CPMG ^{1}H NMR spectra of a PCOS patient's serum prior to and at the end to the 30 months Pio/Flu/Met polytherapy. Clear differences can be observed between the same serum sample prior to and after the 30 months treatment. The polytherapy induced lipoprotein rearrangements reflected in resonances attributable to both methylene (δ 1.25 ppm) and methyl (δ 0.85 ppm) terminal groups of fatty acids contained in LDL and VLDL particles. In contrast, a depletion of the 1,2-propanediol doublet at 1.14 ppm and the acetate broad singlet arising at 1.91 ppm was observed after the treatment. Also, two prominent unidentified resonances arisen in the serum spectra of PCOS patients prior to the treatment were decreased as a result of the polytherapy. The first resonance corresponded to a broad singlet (multiplicity examined by 2D J-res) at 0.85 ppm, and the second unknown resonance was composed of three different peaks centered at 1.18 ppm which probably do not correspond with a

triplet signal because their intensities did not match exactly with the established 1/2/1 triplet intensity ratio.

Table 2 summarizes the statistical values, detailed moieties assignments and structural identities of the metabolites analyzed using CPMG ^{1}H NMR. The relative concentration of glutamate, 1,2-propanediol, lysine, and succinate in serum samples was markedly decreased as a result of the treatment. In contrast, NMR signals attributed to N-$(CH_3)_3$ groups of choline-containing molecules were increased after the treatment. In addition, the three unknown NMR signals centered at 1.18 ppm, which match the characteristic pattern of resonances found in the serum spectra of patients with coronary heart disease (CHD) reported by Jankowski et al. [25], decreased in the spectra after the treatment. Such specific pattern of resonances corresponded to oxidized LDL particles, as demonstrated by Jankowski and co-workers when they compared the spectrum of lipoprotein subfractions of patients with CHD and LDL particles oxidized *in vitro* using Cu^{2+}. To confirm the association of lipoprotein particles with these signals we precipitated serum proteins from untreated patients using cold-acetone and recorded the ^{1}H-NMR spectra on the supernatant. As showed in Figure 2B, the three signals centered at 1.18 ppm disappeared after cold acetone precipitation, suggesting an association with lipoprotein-related particles. Overall, our results suggest that the Pio/Flu/Met polytherapy has a dramatic effect on the lipoprotein profile of PCOS patients.

NMR		
Metabolite	p-value	Mean (% Variation)±SEM
unknown (LDL-ox related structure)	2.9E-03	-35±19
1,2-propanediol	7.3E-04	-65±27
unknown (LDL-ox related structure)	7.3E-03	-24±27
Lysine	4.9E-04	-45±4
Acetates	5.8E-04	-55±6
Glutamate	5.8E-04	-54±4
Succinate	5.8E-04	-42±5
sn-Glycero-phospocholine	1.1E-02	120±22
unknown	7.3E-04	156±41
unknown	7.3E-04	-35±5
unknown (LDL-ox related structure)	5.6E-03	-34±10
GC-MS		
Metabolite	p-value	Mean(%Variation)±SEM
1,2-propanediol	6.89E-03	-77±17
Nonanoic acid	6.89E-03	-68±7
Glutaric acid	4.21E-02	-91±6
Glutamic acid	3.74E-03	-58±9
Azelaic acid	2.11E-03	-96±4
LC-MS		
Metabolite	p-value	Mean(%Variation)±SEM
Caprylic acid (C-8)	1.5E-02	183±13
9/13-HODE	1.5E-02	-86±7

Table 2. Summary of the metabolites found to be significantly varied in either analytical platform after 30 months low-dose Pio-Flu-Met polytherapy. Percentage of variation was calculated for each patient as the area of the spectral region or selected XCMS feature at baseline minus the area of the same feature or spectral region after the treatment relative to the former. Values are expressed as mean 6 SEM. A negative value indicates that levels of the corresponding metabolite resulted significantly decreased with the treatment while positive values indicate a significant increase. p-values correspond to Wilcoxon rank-summed test and FDR correction. Statistical significance was considered for those spectral regions or features having p-corrected values,0.05 and fold changes.2;

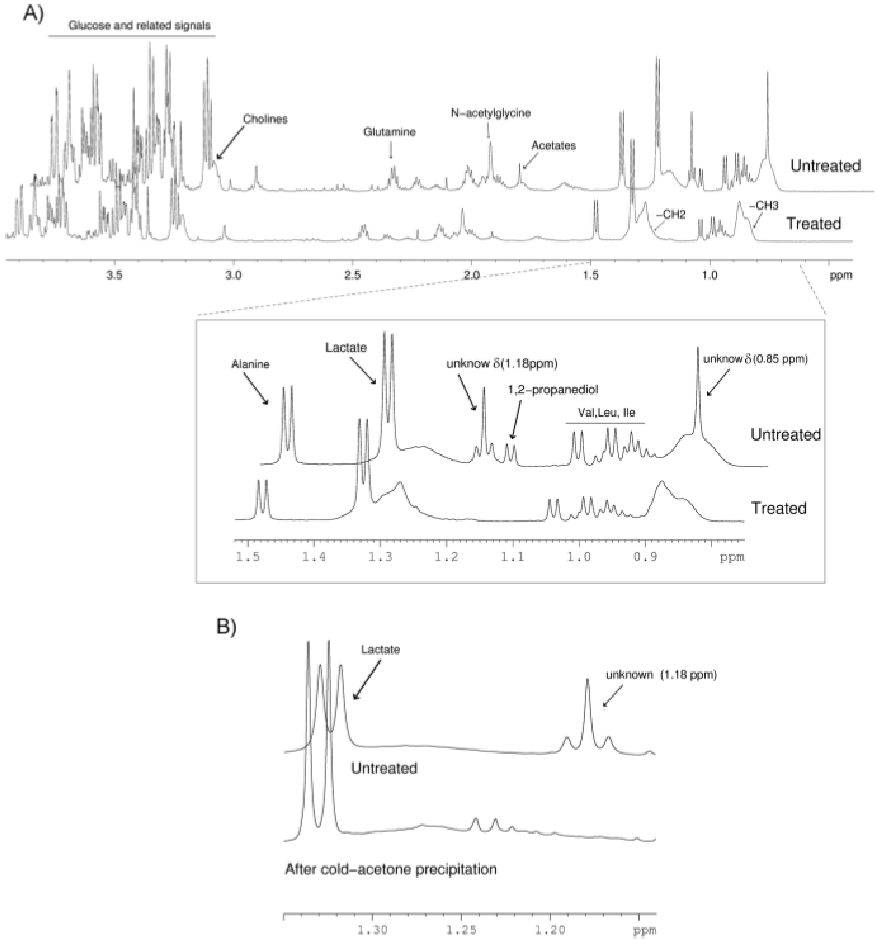


Figure 2. CPMG 1H NMR spectra of a representative PCOS patient's serum. (A) Comparative spectra at baseline and after 30 months of Pio/Flu/Met polytherapy. CPMG spin echo experiment allows filtering broad signals of lipid and lipoproteins enhancing low-molecular weight metabolites such as amino acids, lactate and intermediate metabolites. The inset displays an expanded d (0.75–1.5 ppm) spectral region showing two unidentified resonances characteristic of the serum spectra of untreated PCOS patients: a broad singlet arising at 0.85 ppm and three peaks centered at 1.18 ppm. (B) CPMG 1H-NMR spectra of the same untreated PCOS patient shown in Figure 2A before and after cold acetone precipitation. After acetone precipitation the three signals centered at 1.18 ppm were depleted, confirming the occurrence of oxidized lipoprotein-related structures in the serum of PCOS women.

Lipoprotein rearrangements after low-dose Pio/Flu/Met intervention

To confirm that the lipoprotein profile is severely affected by the Pio/Flu/Met treatment, we measured the size of lipoprotein particles in serum samples prior to and at the end of the 30 months intervention, according to our recently described methodology [21]. In brief, the size of different lipoprotein subclasses were estimated using up to seven Lorentzian functions to fit the methyl peak surface obtained from 2D (bipolar-LED) diffusion edited ^{1}H-NMR experiments. Then, diffusion coefficients and hydrodynamic radius through Stokes-Einstein equation for each one of these seven functions were estimated. Figure 3A shows the average methyl spectrum centered at 0.85 ppm after 30 months of polytherapy in relation to the average signal at baseline. The shift to the upfield region in the spectra after the treatment indicates a greater contribution of the HDL subfraction to the signal of methyl groups. Figure 3B shows the fitted spectrum of a treated PCOS patient using the seven Lorentzian functions. Based on the previously measured diffusion coefficients described in our methodology [21], we estimated the hydrodynamic radius (i.e., size) for each of these functions. Figure 3C shows the mean percentage of variation of the estimated radius for each function as a result of the Pio/Flu/Met treatment. The polytherapy resulted in significantly increased radiuses associated with atherogenic small, dense LDL (F3) and protective large HDL (F4-F6) lipoprotein subclasses.

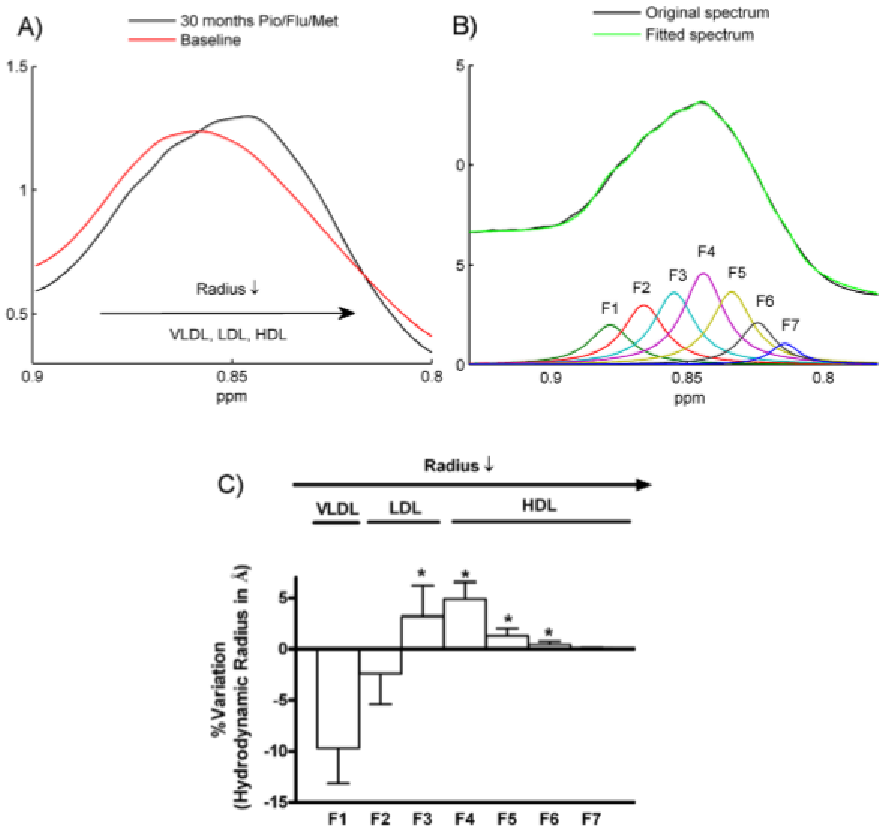


Figure 3. Lipoprotein 1H-NMR analysis. (A) Comparative bipolar-LED diffusion mean spectra of the methyl region (d 0.85 ppm) for untreated and treated PCOS patients. (B) Bipolar LED pulse sequence 1H NMR spectra of a treated PCOS serum showing the fitting of the methyl band using the seven Lorentzian functions derived from our previously described methodology. Black line represents the original methyl envelope and green line thereconstructed spectrum after the fitting. (C) The estimated radiuses of lipoprotein particles in serum calculated using the seven Lorentzian functions were compared at baseline and after 30 months of Pio/Flu/Met polytherapy.

Mass spectrometry analysis

Our mass spectrometry-based platform involves LC-ESI-TOF-MS and GC-single quad MS profiling followed by data analysis with the open-source software XCMS. The relative abundance of metabolites in serum samples was quantified by comparing the

integrated area of each feature, and calculating the percentage of variation of such feature to indicate the level of differential regulation prior to and at the end of the Pio/Flu/Met intervention. Some of the most up-regulated metabolites were identified by tandem MS/MS. Metabolite annotations and statistical analysis are summarized in Table 2. Metabolites identified using GC-MS were consistent with NMR data. For example, the levels of glutamate and 1,2-propanediol in serum were also significantly decreased after the polytherapy. In addition, the treatment led to decreased levels of nonanoic, glutaric and azelaic acid. Of note, the level of azelaic acid resulted positively correlated with carotid IMT ($r=0.92$, $p=3.96\times10^{-7}$) (Figure 4). LC-MS data showed that the treatment resulted in significantly increased level of caprylic acid, whereas it induced a marked reduction in the level of 9-HODE and 13-HODE, the most abundant mono-hydroxyderivative forms resulting from the oxidation of linoleic acid.

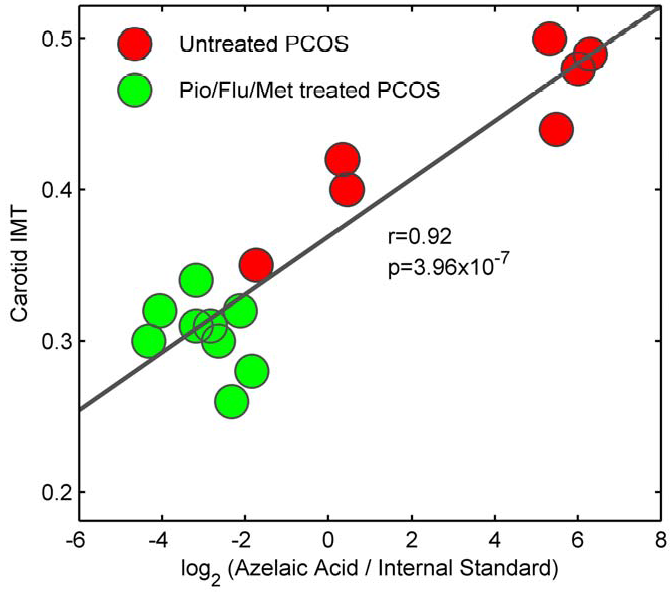


Figure 4. Correlation between IMT and azelaic acid levels in serum. Positive significant correlation (r = 0.92, p = 3.9661027) between carotid IMT values and azelaic acid levels. Azelaic acid levels were calculated as the ratio of the fragmentation peak of azelaic acid at m/z = 317 (retention time = 16.38 min) and the peak area of the internal standard. Red and green dots represent values of untreated and treated PCOS patients respectively.

DISCUSSION

The underlying causes of PCOS are still unknown, and therefore the medical treatment is tailored to the patient's symptoms. Typically, these are lowering insulin level, restoration of fertility and regular menstruation, and treatment of hirsutism and acne. General interventions such as low-dose flutamide, metformin and pioglitazone in combination with an estro-progestagen can be very beneficial because it confers further improvements of the endocrine-metabolic state. Our metabolomic analysis has revealed that the polytherapy induces an increase in the estimated radius of small, dense LDL lipoprotein subclasses together with a reduction

in the level of oxidized LDL particles. These macromolecular rearrangements are associated with changes in the level of specific downstream metabolites produced by linoleic acid peroxidation (i.e. 9-HODE, 13-HODE). Linoleic acid represents the most abundant polyunsaturated fatty acid in LDL particles. Oxidation of LDL transforms linoleic acid into different hydroperoxyderivative (HPODEs) isomers [26], which are subsequently reduced and released by specific lipases from the membrane lipids as free hydroxyoctadecadienoic acid (HODE), such as 9- and 13-HODE, identified in our study [27] (Figure 5). In addition, the oxidative modification and degradation of fatty acids contained in LDL particles generates a complex array of shorter chain-length fragments that covalently modify ε-amino groups of lysine residues of the protein moiety to generate the oxidatively modified LDL particles [28,29,30]. Among these shorter chain-length fragments causing structural modifications of proteins, we have identified azelaic acid and glutaric acid [31]. It is worth mentioning that azelaoyl phosphatidylcholine (azPC) accounts for almost two-thirds of the oxidized phospholipids in oxLDL [32,33]. Also, accumulation of azelaic acid and glutaric acid in plasma of diabetic rat models and type 1 diabetic patients, respectively, has been reported previously [34], Overall, our NMR- and MS-based metabolomics study demonstrate that the Pio/Flu/Met polytherapy reduces the amount of oxidized lipoprotein particles and downstream oxidative metabolites in the serum of PCOS patients.

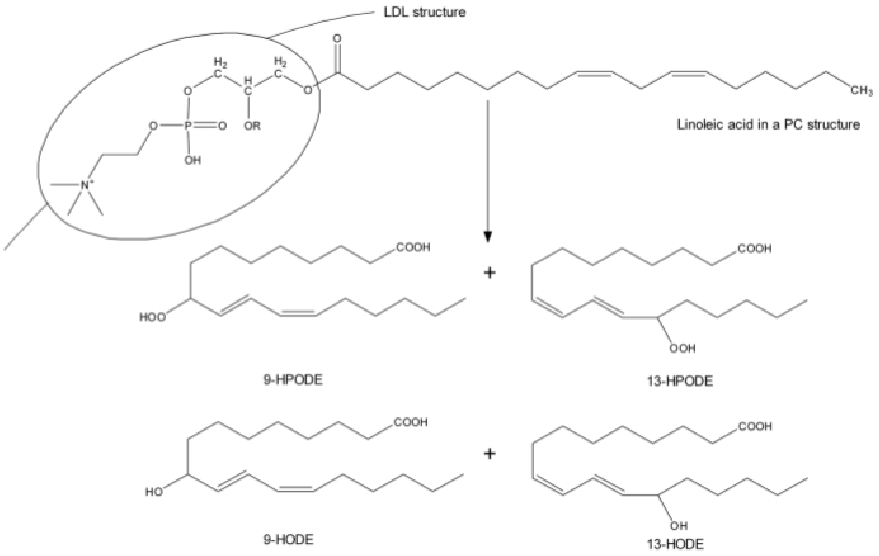


Figure5.Linoleic acid oxidation products.Formation of linoleic acid hydroperoxyderivatives (HPODEs) are further reduced to their corresponding hydroxyderivatives (HODES). Chemical structures of linoleic acid (18:2); 9- and 13-hydroperoxylinoleic acid (9- and 13-HPODE); and 9 and 13-hydroxylinoleic acid (9- and 13-HODE).

Previous studies explored the effect of flutamide, metformin, and pioglitazone in monotherapy or in combination therapy. For example, combined Pio/Met therapy in type 2 diabetic patients improved their specific lipid abnormalities [35,36,37,38]. Pioglitazone, when used both in monotherapy or in combination therapy, modifies the atherogenic lipoprotein profile reducing triglycerides, increasing the larger HDL_2 subfractions and improving the HDL cholesterol load [39]. Very few studies, however, have addressed in detail the effects of pioglitazone and metformin in LDL subfractions. Lawrence et al [40] reported a significant fall in LDL_3 mass and LDL_3 proportion in overweight type 2 diabetic patients treated with metformin alone. The total cholesterol–to–

apoB ratio (used as a surrogate marker for changes in LDL subfraction distribution), however, remained unchanged. In contrast, when overweight type 2 diabetic patients were treated with pioglitazone alone, a significant increase in the cholesterol-to-apoB ratio was reported, indicating larger (and potentially less atherogenic) LDL particles. Pioglitazone alone also induced an increase of LDL particles diameter and a decrease in LDL density in normolipidemic, nondiabetic patients with hypertension [41]. Finally, the treatment of Goto-Kakizaki rats (a type 2 diabetes model) with pioglitazone reduced the levels of lipid peroxides in plasma and the susceptibility of LDL particles to oxidation [42]. Our findings result in good agreement with previous studies and evidence that the addition of pioglitazone to the combined flutamide/metformin polytherapy induces an increase in the mean diameter particles of small LDL subfractions. Besides, our comprehensive metabolomic approach allowed us not only to detect changes in the size of the different lipoprotein particles but also in downstream, oxidation products such as 9- and 13-HODE, azelaic acid and glutaric acid. Overall, our results demonstrate the utility of metabolomics to explore the effect of medical treatments on metabolic alterations. In this study, the combined pioglitazone/flutamide/metformin polytherapy reverses the oxidant status of untreated PCOS patients. Given that oxidation of LDL particles has been suggested to be the key triggering event in the progression of atherosclerotic lesions [43], and that the carotid IMT is also markedly reduced after 30 months of polytherapy, we

postulate that untreated young PCOS women may suffer early stages of atherosclerosis that could potentially have deleterious effects at older ages. Besides, we have demonstrated that azelaic acid levels are strongly correlated with carotid IMT. Hence we suggest that azelaic acid can be considered as an early marker of lipoprotein oxidation and subclinical atherosclerosis.

REFERENCES

1. Ibanez L, Valls C, de Zegher F (2006) Discontinuous low-dose flutamide-metformin plus an oral or a transdermal contraceptive in patients with hyperinsulinaemic hyperandrogenism: normalizing effects on CRP, TNF-alpha and the neutrophil/lymphocyte ratio. Human Reproduction 21: 451-456.

2. Lambrinoudaki I (2011) Cardiovascular risk in postmenopausal women with the polycystic ovary syndrome. Maturitas 68: 13-16.

3. McGowan M (2011) Polycystic Ovary Syndrome: A Common Endocrine Disorder and Risk Factor for Vascular Disease. Current Treatment Options in Cardiovascular Medicine: 1-13.

4. Macut D, Panidis D, Glisić B, Spanos N, Petakov M (2008) Lipid and lipoprotein profile in women with polycystic ovary syndrome. Can J Physiol Pharmacol 86: 199-204.

5. Macut D, Damjanovic S, Panidis D, Spanos N, Glisic B, et al. (2006) Oxidised low-density lipoprotein concentration - early marker of an altered lipid metabolism in young women with PCOS. Eur J Endocrinol 155: 131-136.

6. Sabuncu T, Vural H, Harma M, Harma M (2001) Oxidative stress in polycystic ovary syndrome and its contribution to the risk of cardiovascular disease. Clinical Biochemistry 34: 407-413.

7. Wild RA (2002) Long-term health consequences of PCOS. Human Reproduction Update 8: 231-241.

8. Legro RS (2003) Polycystic Ovary Syndrome and Cardiovascular Disease: A Premature Association? Endocrine Reviews 24: 302-312.

9. Talbott EO, Zborowski JV, Rager JR, Boudreaux MY, Edmundowicz DA, et al. (2004) Evidence for an association between metabolic cardiovascular syndrome and coronary and aortic calcification among women with polycystic ovary syndrome. J Clin Endocrinol Metab 89: 5454-5461.

10. Ibanez L, Lopez-Bermejo A, del Rio L, Enriquez G, Valls C, et al. (2007) Combined low-dose pioglitazone, flutamide, and metformin for women with androgen excess. Journal Of Clinical Endocrinology And Metabolism 92: 1710-1714.

11. Ibanez L, Lopez-Bermejo A, Diaz M, Enriquez G, Valls C, et al. (2008) Pioglitazone (7.5 mg/day) added to flutamide-metformin in women with androgen excess: additional increments of visfatin and high molecular weight adiponectin. Clinical Endocrinology 68: 317-320.

12. Ibáñez L, López-Bermejo A, Díaz M, Enríquez G, Del Río L, et al. (2009) Low-dose pioglitazone and low-dose flutamide added to metformin and estro-progestagens for hyperinsulinemic women

with androgen excess: add-on benefits disclosed by a randomized double-placebo study over 24 months. Clin Endocrinol 71: 351-357.

13. Ibañez L, López-Bermejo A, Díaz M, Enríquez G, Del Río L, et al. (2010) Low-dose pioglitazone, flutamide, metformin plus an estro-progestagen for non-obese young women with polycystic ovary syndrome: increasing efficacy and persistent safety over 30 months. Gynecological Endocrinology 26: 869-873.

14. Lenz EM, Bright J, Knight R, Wilson ID, Major H (2004) Cyclosporin A-induced changes in endogenous metabolites in rat urine: A metabonomic investigation using high field 1H NMR spectroscopy, HPLC-TOF/MS and chemometrics. Journal of Pharmaceutical and Biomedical Analysis 35: 599-608.

15. Sieber M, Wagner S, Rached E, Amberg A, Mally A, et al. (2009) Metabonomic study of ochratoxin a toxicity in rats after repeated administration: phenotypic anchoring enhances the ability for biomarker discovery. Chem Res Toxicol 22: 1221-1231.

16. Williams R, Lenz EM, Wilson AJ, Granger J, Wilson ID, et al. (2006) A multi-analytical platform approach to the metabonomic analysis of plasma from normal and zucker (fa/fa) obese rats. Molecular BioSystems 2: 174-183.

17. Dunn WB, Bailey NJC, Johnson HE (2005) Measuring the metabolome: current analytical technologies. Analyst 130: 606-625.

18. Moco S, Vervoort J, Moco S, Bino RJ, De Vos RCH, et al. (2007) Metabolomics technologies and metabolite identification. TrAC Trends in Analytical Chemistry 26: 855-866.

19. Lenz EM, Wilson ID (2006) Analytical Strategies in Metabonomics. Journal of Proteome Research 6: 443-458.

20. Dunn WB, Ellis DI (2005) Metabolomics: Current analytical platforms and methodologies. TrAC Trends in Analytical Chemistry 24: 285-294.

21. Mallol R, Rodríguez M, Heras M, Vinaixa M, Cañellas N, et al. (2011) Surface fitting of 2D diffusion-edited 1H NMR spectroscopy data for human plasma lipoproteins characterization. Metabolomics: 1-11.

22. Palazoglu M, Fiehn O (2009) Metabolite Identification in Blood Plasma Using GC-MS and the Agilent Fiehn GC-MS Metabolomics RTL Library. Agilent Application Note.

23. Jansen JJ, Hoefsloot HCJ, van der Greef J, Timmerman ME, Smilde AK (2005) Multilevel component analysis of time-resolved metabolic fingerprinting data. Analytica Chimica Acta 530: 173-183.

24. Smith CA, Want EJ, O'Maille G, Abagyan R, Siuzdak G (2006) XCMS: Processing Mass Spectrometry Data for Metabolite Profiling Using Nonlinear Peak Alignment, Matching, and Identification. Analytical Chemistry 78: 779-787.

25. Jankowski J, Nofer JR, Tepel M, Griewel B, Schläter H, et al. (1998) Identification of oxidized low-density lipoprotein in human serum by NMR spectroscopy. Clin Sci 95: 489-495.

26. Lee SH, Blair IA (2001) Oxidative DNA Damage and Cardiovascular Disease. Trends in Cardiovascular Medicine 11: 148-155.

27. Lee SH, Williams MV, Blair IA (2005) Targeted chiral lipidomics analysis. Prostaglandins & Other Lipid Mediators 77: 141-157.

28. Palinski W, Rosenfeld ME, Ylä-Herttuala S, Gurtner GC, Socher SS, et al. (1989) Low density lipoprotein undergoes oxidative modification in vivo. Proceedings of the National Academy of Sciences of the United States of America 86: 1372-1376.

29. Haberland ME, Olch CL, Folgelman AM (1984) Role of lysines in mediating interaction of modified low density lipoproteins with the scavenger receptor of human monocyte macrophages. Journal of Biological Chemistry 259: 11305-11311.

30. Raghavamenon A, Garelnabi M, Babu S, Aldrich A, Litvinov D, et al. (2009) Alpha-Tocopherol Is Ineffective in Preventing the Decomposition of Preformed Lipid Peroxides and May Promote the Accumulation of Toxic Aldehydes: A Potential Explanation for the Failure of Antioxidants to Affect Human Atherosclerosis. Antioxidants & Redox Signaling 11: 1237-1248.

31. Januszewski AS, Alderson NL, Jenkins AJ, Thorpe SR, Baynes JW (2005) Chemical modification of proteins during peroxidation of phospholipids. J Lipid Res 46: 1440-1449.

32. Tokumura A, Toujima M, Yoshioka Y, Fukuzawa K (1996) Lipid peroxidation in low density lipoproteins from human plasma and egg yolk promotes accumulation of 1-acyl analogues of platelet-activating factor-like lipids. Lipids 31: 1251-1258.

33. Chen R, Yang L, McIntyre TM (2007) Cytotoxic Phospholipid Oxidation Products. Journal of Biological Chemistry 282: 24842-24850.

34. Januszewski AS, Jenkins AJ, Baynes JW, Thorpe SR (2005) Lipid-Derived Modifications of Plasma Proteins in Experimental and Human Diabetes. Annals of the New York Academy of Sciences 1043: 404-412.

35. Einhorn D, Rendell M, Rosenzweig J, Egan JW, Mathisen AL, et al. (2000) Pioglitazone hydrochloride in combination with metformin in the treatment of type 2 diabetes mellitus: A randomized, placebo-controlled study. Clinical Therapeutics 22: 1395-1409.

36. Charbonnel B, Schernthaner G, Brunetti P, Matthews D, Urquhart R, et al. (2005) Long-term efficacy and tolerability of add-on pioglitazone therapy to failing monotherapy compared with addition of gliclazide or metformin in patients with type 2 diabetes. Diabetologia 48: 1093-1104.

37. Derosa G, D'Angelo A, Ragonesi PD, Ciccarelli L, Piccinni MN, et al. (2006) Metformin–pioglitazone and metformin–rosiglitazone effects on non-conventional cardiovascular risk factors plasma level in type 2 diabetic patients with metabolic syndrome. Journal of Clinical Pharmacy and Therapeutics 31: 375-383.

38. Derosa G, D'Angelo A, Ragonesi PD, Ciccarelli L, Piccinni MN, et al. (2007) Metabolic effects of pioglitazone and

rosiglitazone in patients with diabetes and metabolic syndrome treated with metformin. Internal Medicine Journal 37: 79-86.

39. Hanefeld M (2009) The role of pioglitazone in modifying the atherogenic lipoprotein profile. Diabetes, Obesity and Metabolism 11: 742-756.

40. Lawrence JM, Reid J, Taylor GJ, Stirling C, Reckless JPD (2004) Favorable Effects of Pioglitazone and Metformin Compared With Gliclazide on Lipoprotein Subfractions in Overweight Patients With Early Type 2 Diabetes. Diabetes Care 27: 41-46.

41. Winkler K, Konrad T, Füllert S, Friedrich I, Destani R, et al. (2003) Pioglitazone Reduces Atherogenic Dense LDL Particles in Nondiabetic Patients With Arterial Hypertension. Diabetes Care 26: 2588-2594.

42. Iida KT, Kawakami Y, Suzuki M, Shimano H, Toyoshima H, et al. (2003) Effect of thiazolidinediones and metformin on LDL oxidation and aortic endothelium relaxation in diabetic GK rats. American Journal of Physiology - Endocrinology And Metabolism 284: E1125-E1130.

43. Steinberg D (1990) Arterial Metabolism of Lipoproteins in Relation to Atherogenesis. Annals of the New York Academy of Sciences 598: 125-135.

CHAPTER 3. DISCUSSION

The ultimate goal of untargeted metabolomics is the comprehensive characterization and quantitation of small molecule in biological matrices. However, due to large chemical diversity of metabolites (e.g., sugars, aminoacids, vitamins, hormones, lipids, etc.), the achievement of this goal will depend largely on both the analytical platform and the sample extraction protocol used. And in this sense, there is not a unique sample preparation protocol neither an analytical platform that allow complete metabolome coverage.

Therefore, it is desirable to implement different methodological procedures, including various metabolite extractions, combination of MS ion sources [24] , different type of chromatographic column [9] and more than one analytical platforms (NMR or MS).

In practice, however, the vast majority of metabolomic studies have been performed by using a single analytical platform based on either NMR [116,117] or MS [16,118] technology. Few studies have taken full advantage of the complementarity between NMR and MS [12,17]. Apart from the need of experitse in MS and NMR and the availability of both technologies, the main reason for using a single platform is the disparity between sample preparation for NMR and MS analysis. Therefore, in the few studies that have aimed to measure the same sample by using LC-MS and NMR, the original sample was aliquoted and subjected to different extraction protocols [119]. Alternatively, the supernatant obtained after the extraction process is aliquoted, dried an later dissolved in specific

solvents for NMR or LC-MS [120]. This common practice requieres initial large sample volumes, which may preclude the analysis of important biofluids, such as vitrous humor or cerebral spinal fluid (CSF).

In this context, section 2.1.1 describes the optimization of a metabolite extraction and sample preparation protocol for measuring metabolites by NMR followed by LC-MS without further pre-treatment. Twelve extraction methods were tested, using a combination of different solvents and temperatures. The results of this study showed that the choice of solvents largely determined extraction efficiencies, and extraction methods based on methanol/chloroform/water (7:2:1) and methanol/water (1:1) best showed the complementarity of NMR and LC-MS in metabolomics, providing high extraction yields and minimal repetition of metabolites between platforms. However, the most striking result derived from this study was the demostration of deuterated solvents used for NMR have essentially no effect on the MS data. This opens up a new opportunities to explore compatible extraction procedures for both NMR and MS metabolomics experiments with the aim to increase the metabolome coverage using one single sample aliquote. The study also shows the complementarity of NMR and LC-MS allowing those metabolites that are not readily ionized by ESI such as glucose, or metabolites that are not retained using reverse-phase chromatography such as lactate, can be detected directly using NMR technology. At the same time, metabolites at very low concentration such as

pantothenic acid or acetil CoA are not detected by NMR and, instead, are detected by LC. This complementarity between analytical platforms should gather more information about the levels of metabolites perturbed, thus enabling a wider interrogation of the fundamental biochemical processes involved in the underlying phenotype.

After sample preparation and data adquisition by NMR or MS, the next critical step in our untargeted metabolomics workflow is data pre-processing and data analysis. Metabolomics data analysis aims to identify metabolite features that are differentially altered between sample groups. Metabolomic technologies, however, produce large amounts of data (particularly MS platforms), and dealing with such complex datasets can exert a significant influence on the quality of metabolite quantification and identification. Unfortunately, a standarized protocol for data and statistical analysis in metabolomics has not been stablished.

To partially address this issue, section 2.1.2 of this thesis introduces a robust data analysis workflow. This workflow is based on XCMS software, wich provide a method for peak picking, retention time alignment and quantification of metabolite features in MS metabolomics experiments. Its important to remind that this software does not output metabolite identifications. Rather, it provides a table containing thousand of features with their maximum intensity (or integrated area) for each of them in each sample. Determining the identity of thousands of features is

virutally impossible. For this reason, it is imperative to implement filters to reduce the complexity of the data.

Our first filter deals with analytical variation: most common sources of analytical variation in LC-MS and GC-MS experiments results from sample preparation, instrumental drifts caused by chromatographic columns and MS detectors, and errors caused in data processing [99]. Here, analytical variation is handled by using quality controls (QC) [121]. QC are pools of individual samples analyzed in the experiment, which are analyzed periodically throughout the sample worklist. The variation observed in QC samples can be calculated for each feature and data preprocessing algorithms can be applied to remove those features that show larger analytical than biological variation.

Next, statistical analysis is performed. This step is pivotal for untargeted metabolomic experiments, since only those features showing statistically significant differences among groups will be taken for MS/MS identification [72,122]. The majority of metabolomic studies use multivariate models to report their main findings. However, untargeted metabolomic datasets can also be approached from a univariate perspective using traditional statistical methods that consider only one variable at a time. The use of both methodologies maximizes the extraction of relevant information from metabolomics datasets.

In this thesis, the impact of univariate statistical analysis on LC-MS based metabolomic experiments was studied in section 2.1.2. We aimed to stablish a practical guide with resources to overcome

common challenges when performing multiple univariate analysis. The ultimate goal was to constraint the number of initially detected features to an amenable number for further MS/MS identification experiments. Hence, the choice of potential metabolites is purely an statistical decision. In summary, section 2.1.2 demonstrates the importance of QC, the choice of the best univariate test based on a specific experimental design, the correction of p-values by implementing false discovery rate tests, and the use of fold changes to numerically represent differences among groups.

Finally, those features passing our statistical criteria are searched in dataases such as Metlin or HMDB. Features that return a hit in databases are considered putative metabolites, which will be further verified by MS/MS and/or retention time using pure standards. Given that this represent the rate-limiting step of the untargeted metabolomic workflow, it is essential to reduce and prioritize features from raw data.

Real untargeted metabolomics examples are shown in section 2.2, where xcms software [56] has been used for GC and LC data pre-processing. In all cases, feautres derived from statistical analysis are initially identified as putative metabolites using different libraries such as Nist or Fiehn for GC and Metlin for LC. Metabolites are unequivocally confirmed by matching retention time and MS/MS spectra with that of pure chemical standards. Despide this high degree of confirmation, unamiguous assignments

of observed metabolite features to a single metaboliet are not always achievable.

Finally, identified metabolites must be interpreted in a biological/biochemical context. In this thesis, women with PCOS have been studied, and our metabolomic results are discussed on the basis of this physio-pathological condition. PCOS is associated with long-term risk factors for the development of severe metabolic disorders including obesity, diabetes and cardiovascular disease [87]. The pathogenesis of PCOS is complex and its aetiology remains uncelar. This complexity presents challenges for a full understanding of the molecular pathways that contribute to the development of this major disease. In this thesis, key metabolomic methodologies, including NMR and MS, have been used to provide new mechanistics insights on PCOS.

Section 2.2.1 describes two metabolomic studies aimed to discover clinically relevant pathways that are affected by PCOS: (i) in the first work (section 2.2.1.1), we studied the impact that obesity exerts on the metabolic disorder of women with PCOS. This was studied by applying an untargeted GC-MS-based metabolomics approach to plasma samples. In this work, metabolomic signatures for obese PCOS patients were compared with signatures for lean PCOS patients and control women, with the aim to elucidate how abdominal adiposity and obesity affect PCOS. Approximately 50% of women with PCOS are overweight or obese and most of them have abdominal obesity. Obesity, and particularly abdominal obesity, may be partly responsible for

insulin resistance and associated hyperinsulinemia in women with PCOS.

Our results show that obese PCOS women present elevated levels of glycerol and plasma long-fatty acids including linoleic and oleic acid, which suggest increased lypolysis in these patients. Conversely, lean PCOS women showed a metabolic profile characterized by reduced lipolysis and increased glucose utilization (i.e., increased lactic acid concentration) in peripheral tissues. Furthermore, women with PCOS were associated with decreased α-tocopherol levels independently of obesity. These results show that substantial metabolic heterogeneity, strongly influenced by obesity, underlies PCOS. Moreover, both obese and lean patients with PCOS were hyperinsulinemic and insulin resistant compared with the control group. We postulate that hyperinsulineamia may occur in the absence of insulin resistance in some women with PCOS, particularly in those without abdominal adiposity and/or obesity. Consequently, we decided to focus our PCOS research in young lean PCOS patients, understood as an homogeneous group of patients with similar phenotypes.

In the second work (section 2.2.1.2), we elucidated new metabolic insights on PCOS in a cohort of young lean PCOS patients by applying a untargeted metabolomic approach based on NMR and LC-MS to serum samples.

Our results indicate that women with PCOS activate the biosynthesis of GSH through the γ-glutamyl cycle to compensate their oxidative imbalance (i.e., redox status). This is reflected in

greater levels of GSH (and GSH/GSSG ratio) in PCOS. In addition, our NMR-based lipoprotein analysis showed a dyslipidemic profile in women with PCOS that we also associate with an atherogenic profile in these women. Previous studies described that oxidative stress can cause lipoprotein modifications. Our metabolomic results revealed alterations in the levels of methionine and methionine sulfoxide in PCOS women, which we postulated was indicative of increased oxidation of methionine residues in apo-A1 from HDL particles. We confirmed our hypothesis by measuring the the amount of peptide containing methionine-148 residue by MALDI-TOF MS, and calculating the ratio of oxidized Met-148/Met-148 in apo-A1 in control and PCOS women. Overall, the combination of metabolomics and proteomics revealed that the atherogenic profile of women with PCOS could result from an elevated oxidative stress that, in turn, impact HDL functionality through oxidation of apo-A1. Therefore, we propose that elevated levels of methionine sulfoxide in serum, oxidized methionine-148 in apo-A1, and the lipoprotein profile by NMR might constitute earlier biomarkers of metabolic syndrome.

Another potential application of metabolomics is the study of drug response phenotypes. This type of studies provide possible mechanisms and biomarkers responsible for variation in drug response. Therefore, the global mapping of metabolite signatures pre- and post drug treatment could reveal new metabolic responses to drug treatment. In this context, the third work of this thesis (section 2.2.2) illustrates the potential of metabolomics to

elucidate novel biochemical responses to polytherapy in women with PCOS. This study benefited from the implementation of three different analytical platforms (i.e., NMR, LC-MS and GC-MS) to the study of serum PCOS samples before and after low-dose metformin, flutamide and pioglitazone in combination with an estroprogestagen. This new polytherapy decreased levels of oxidized LDL particles in serum, as well as downstream metabolic oxidation products of LDL particles such as 9- and 13-HODE, azelaic acid and glutaric acid. In contrast, the polytherapy increased the size of small LDL and large HDL particles. Clinical and endocrine-metabolic markers were also monitored, showing that the level of HDL cholesterol was increased after the treatment, whereas the level of androgens and the carotid intima-media thickness were reduced. Significantly, the abundance of azelaic acid and the carotid intima-media thickness resulted in a high degree of correlation. Altogether, our results reveal that this polytherapy markedly reverts the oxidant status of untreated PCOS women, and potentially improves the pro-atherogenic condition in these patients.

In summary, our metabolomic studies converge in a new hypothesis for PCOS (Figure 5) and, ultimately, for diabetes: hyperandrogenism, understood as high serum testosterone levels, and insuline resistance are common traits in women with PCOS. In turn, insulin resistance is thougt to promote the hyperandrogenism through the compensatory hyperinsulinemia [123]. Importnantly, our PCOS patients present hyperandrogenism, oxidative stress and

a dyslipidemic profile, without hyperglycemia. We showed that dyslipidemic profile result, in part, to lipoprotein modifications induced by oxidative stress. Several characteristics such as androgen excess, abdominal adiposity, insulin resistance and obesity in PCOS patients may contribute to the development of local and systemic oxidative stress which may reciprocally worsen these metabolic abnormalities [82]. However, our young lean PCOS patients are solely characterized by androgen excess. Recently, a couple of studies have demonstrated that hyperandrogenemia and glucose ingestion induce oxidative stress [123,124]. Consequently, we postulate that hyperandrogenism in combination with glucose intake induce oxidative stress and ROS production at early stages in young lean women with PCOS. This will promote the oxidation of lipoprotein particles and ultimately will result in dyslipidemia. Therefore, the dyslipidemic profile predates any other complication associated with PCOS such as inuslin resistance or type 2 diabetes. More studies are needed to confirm our mechanism, however, it seems evident that this hypothesis can be extrapolated to pre-diabetic patients opening up new possibilities for diagnostic.

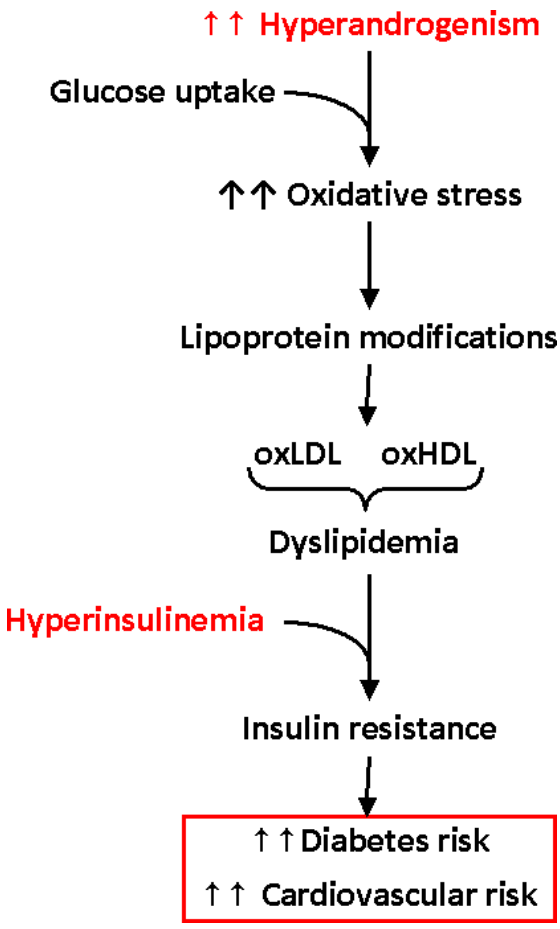


Figure 5. Schematic representation of hypothesis for PCOS

Then, results presented in this thesis demonstrated the value of metabolomics as a global profiling tool in discovery-driven of metabolic dysregulation pathways in PCOS disease. All these works together provide novel biochemical insight into PCOS disease and have opened a window into the metabolic dysregulation of PCOS disease.

This thesis demonstrates the utility of metabolomics for global profiling studies. In particular, the clinical studies here presented have revealed novel biochemical insights into the metabolic

dysregulation of women with PCOS. Furthermore, our serum metabolomic studies open up new opportunities for diagnostics and can facilitate monitoring of disease progression and therapeutic intervention. I anticipate that human diagnosis based on single molecule biomarkers will be replaced in the near future by metabolomics based multiparameter diagnosis, and may represent an extremely promising advance toward early detection of diseases such as PCOS.

CHAPTER 4. CONCLUSIONS

WORK 1

- The choice of solvents largely determined extraction efficiencies.
- Methanol-containing methods are more effective in protein removal than those containing acetonitrile
- Methanol/chloroform/water (7/2/1) and methanol/water (1/1) provide higher extraction yields and minimal repetition of metabolites, showing the complementarity of NMR and LC-MS in metabolomics.
- Deuterated solvents used for NMR have essentially no effect on the MS data.
- The choice of solvent had more influence than the extraction temperature.

WORK 2

- Statistics is an essential tool for choosing those metabolites that will be identificate in LC-MS/MS.
- The use of QC is an important issue in untargeted metabolomics experiments for handling analytical variation.
- Metabolomics data analysis can be approached from a univariate perspective using traditional statistical methods.
- A wide variety of univariate statistical tests to compare mean or medians are available for untargeted metabolomics experiments.

- Experimental design and data distribution are the two key considerations for choosing the correct univariate statistical test.
- Multiple univariate statistical test is innevitably linked to false positives rates. Therefore, p-values must be corrected by false discovery tests.
- A robust data analysis workflow for LC-MS represents the first step for a successful identification of relevant metabolites involved in biological phenomena.

WORK 3

- Untargeted metabolomics approach is a powerful tool for the study of PCOS.
- The possibility that hyperinsulinemia may occur in the absence of universal insulin resistance in some women with PCOS, especially when abdominal adiposity and/or obesity are not marked, should be considered when designing diagnostic and therapeutic strategies for the management of this prevalent disorder.

WORK 4

- Metabolomics is a powerful tool to study new mechanistic insights in young lean PCOS patients.
- Untargeted metabolomic approach has revealed that gamma-glytamyl cycle and oxidative stress is involved in PCOS pathogenicity.
- Elevated levels of methionine sulfoxide in serum, oxidized methionine-148 in apo-AI, and the lipoprotein

profile might constitute earlier biomarkers of metabolic syndrome.

WORK 5

- Metabolomics is a powerful technique that permits exploring the effect of medical treatments on metabolic alterations.
- The use of untargeted metabolomic approach has allowed not only to detect changes in the size of different lipoprotien particles but also in downstream, oxidation products such as 9 and 13- HODE, azelaic and glutaric acid.
- The combined pioglitazone/flutamide/metformin polytherapy reverses the oxidant status of untreated PCOS patients.

As a general conclusion, metabolomics can be considered a powerful tool for the study of metabolic disorders. Furthermore, metabolite profiling has demonstrated feasibility and flexibility for revealing new mechanistic insights in metabolic disorders that are not been consider when classical analysis is used. Therefore, our metabolomic analysis have demonstrated a great potential as a useful diagnostic technique and can facilitate monitoring of both disease progression and effects of therapeutic treatment. However, further efforts to understand the diversity of inputs to the metabolome will be an important resource for future metabolomic studies.

CHAPTER 5. REFERENCES

1. Monteiro M, Caravalho M, Bastos M, Guedes de Pinho P Metabolomics Analysis for Biomarker Discovery: Advances and Challenges.
2. Patti GJ, Yanes O, Siuzdak G (2012) Metabolomics: the apogee of the omics trilogy. Nature Reviews Molecular Cell Biology 13: 263-269.
3. Griffiths W, Koal T, Wang Y, Kohl M, Enot D, et al. (2010) Targeted metabolomics for biomarker discovery. Angew Chem Int Ed Engl 49: 5426-5445.
4. Han J, Tschernutter V, Vera Y (2013) Analysis of Selected Sugars and Sugar Phosphates in Mouse Heart Tissue by Reductive Amination and Liquid Chromatography-Electrospray Ionization Mass Spectrometry. ANALYTICAL CHEMISTRY 85: 5965-5973.
5. Yuan W, Anderson KW, Li S, Edwards JL Subsecond absolute quantitation of amine metabolites using isobaric tags for discovery of pathway activation in mammalian cells. Anal Chem 84: 2892-2899.
6. Galuska C, Hartmann M, Sanchez-Guijo A (2013) Profiling intact steroid sulfates and unconjugated steroids in biological fluids by liquid chromatography-tandem mass spectrometry (LC-MS-MS). ANALYST 138: 3792-3801
7. Bruins M, Dane A, Strassburg K (2013) Plasma oxylipin profiling identifies polyunsaturated vicinal diols as responsive to arachidonic acid and docosahexaenoic acid intake in growing piglets. JOURNAL OF LIPID RESEARCH 54: 1598-1607.
8. Putri S, Yamamoto S, Tsugawa H, Fukusaki E (2013) Current metabolomics: Technological advances. Journal of Bioscience and Bioengineering 116: 9-16.
9. Yanes O, Tautenhahn R, Patti GJ, Siuzdak G Expanding Coverage of the Metabolome for Global Metabolite Profiling. Analytical Chemistry 83: 2152-2161.
10. Vuckovic D (2012) Current trends and challenges in sample preparation for global metabolomics using liquid chromatography – mass spectrometry. Anal Bional Chem 403: 1523-1548.

11. Prasad R, Ferenci T (2003) Global metabolite analysis: the influence of extraction methodology on metabolome profiles of Escherichia coli. Analytical Biochemistry 313: 145-154.
12. Brugnara L, Vinaixa M, Murillo S, Samino S, Angel Rodriguez M, et al. (2012) Metabolomics Approach for Analyzing the Effects of Exercise in Subjects with Type 1 Diabetes Mellitus. Plos One 7.
13. Rammouz R, Letisse F, Durand S, Portais J, Wadih Z, et al. (2010) Analysis of skeletal muscle metabolome: Evaluation of extraction methods for targeted metabolite quantification using liquid chromatography tandem mass spectrometry. Analytical Biochemistry 398: 169-177.
14. Roessner U, Nahid A, Chapman B, Hunter A, Bellgard M (2011) Metabolomics - the combination of analytical chemistry, biology and informatics. New York, United States: Elsevier Science pp. 447-459.
15. Nemes P, Rubakhin SS, Aerts JT, Sweedler JV (2013) Qualitative and quantitative metabolomic investigation of single neurons by capillary electrophoresis electrospray ionization mass spectrometry. Nature Protocols 8: 783-799.
16. Emond P, Mavel S, Aidoud N, Nadal-Desbarats L, Montigny F, et al. (2013) GC-MS-based urine metabolic profiling of autism spectrum disorders. Analytical and Bioanalytical Chemistry 405: 5291-5300.
17. Jia H-m, Feng Y-f, Liu Y-t, Chang X, Chen L, et al. (2013) Integration of H-1 NMR and UPLC-Q-TOF/MS for a Comprehensive Urinary Metabonomics Study on a Rat Model of Depression Induced by Chronic Unpredictable Mild Stress. Plos One 8.
18. Dettmer K, Aronov P, Hammock B (2007) MASS SPECTROMETRY-BASED METABOLOMICS. Mass Spectrometry Review 26: 51-78.
19. Zuojing L, Qing L, Lulu G (2013) Use of the Local False Discovery Rate for Identification of Metabolic Biomarkers in Rat Urine Following Genkwa Flos-Induced Hepatotoxicity. Plos One 8.
20. Barri T, Dragsted L (2013) UPLC-ESI-QTOF/MS and multivariate data analysis for blood plasma and serum metabolomics:

Effect of experimental artefacts and anticoagulant. ANALYTICA CHIMICA ACTA 768: 118-128.

21. Ma B, Liu J, Zhang Q, Ying H, Jiye A, et al. (2013) Metabolomic Profiles Delineate Signature Metabolic Shifts during Estrogen Deficiency-Induced Bone Loss in Rat by GC-TOF/MS. Plos One 8.
22. Coulier L, Muilwijk B, Bijlsma S, Noga M, Tienstra M, et al. (2013) Metabolite profiling of small cerebrospinal fluid sample volumes with gas chromatography-mass spectrometry: application to a rat model of multiple sclerosis. Metabolomics 9: 78-87.
23. Pasikanti K, Ho P, Chan E (2008) Gas chromatography/mass spectrometry in metabolic profiling of biological fluids. Journal of Chromatography B 871: 202-211.
24. Warren C (2013) Use of chemical ionization for GC-MS metabolite profiling. Metabolomics 9: 110-120.
25. Wen T, Gao L, Wen Z (2013) Exploratory investigation of plasma metabolomics in human lung adenocarcinoma. Molecular bioSystems 9: 2370-2378.
26. Tsugawa H, Bamba T, Shinohara M, Nishiumi S, Yoshida M, et al. (2011) Practical non-targeted gas chromatography/mass spectrometry-based metabolomics platform for metabolic phenotype analysis Journal of Bioscience and Bioengineering 112: 292-298.
27. Zhang F, Yu C, Wang W (2012) Rapid simultaneous screening and identification of multiple pesticide residues in vegetables. ANALYTICA CHIMICA ACTA 757: 39-47.
28. Rijk J, Peijnenburg A, Blokland M (2012) Screening for Modulatory Effects on Steroidogenesis Using the Human H295R Adrenocortical Cell Line: A Metabolomics Approach Chemical Research in Toxicology 25: 1720-1731.
29. Garcia A, Barbas C (2011) Gas Chromatography-Mass Spectrometry (GC-MS)-Based Metabolomics. 191-204 p.
30. Forcisi S, Moritz F, Kanawati B, Dimitrios T, Lehmann R, et al. (2013) Liquid chromatography–mass spectrometry in metabolomics research: Mass analyzers in ultra high pressure liquid chromatography coupling. Journal of Chromatography A 1292: 51-65.

31. Cubbon S, Antonio C, Wilson J, Thomas-Oates J (2010) Metabolomic applications of HILIC-LC-MS. Mass Spectrometry Review 29: 671-685.
32. Spagou K, Tsoukali H, Raikos N, Gika H, Wilson ID, et al. (2010) Hydrophilic interaction chromatography coupled to MS for metabonomic/metabolomic studies. Journal of separation science 33: 716-727.
33. Patti GJ (2011) Separation strategies for untargeted metabolomics. Journal of separation science 34: 3460-3469.
34. Zhuoling A, Yanhua C, Zhang R (2010) Integrated Ionization Approach for RRLC-MS/MS-based Metabonomics: Finding Potential Biomarkers for Lung Cancer. Journal of Proteome Research 9: 4071-4081
35. Sangeetha R, Bhaskar N, Sounder D (2010) Bioavailability and metabolism of fucoxanthin in rats: structural characterization of metabolites by LC-MS (APCI). Molecular and Cellular Biochemistry 333: 299-310.
36. Ho C, Lam C, Chan M, Cheung R, Law L (2003) Electrospray Ionisation Mass Spectrometry: Principles and Clinical Applications. Clin Biochem Rev 24.
37. Ferreiro-Vera C, Priego-Capote F, Calderon-Santiago M (2013) Global metabolomic profiling of human serum from obese individuals by liquid chromatography-time-of-flight/mass spectrometry to evaluate the intake of breakfasts prepared with heated edible oils. Food Chemistry 141: 1722-1731.
38. Li J, Hoene M, Zhao X (2013) Stable lsotope-Assisted Lipidomics Combined with Nontargeted lsotopomer Filtering, a Tool to Unravel the Complex Dynamics of Lipid Metabolism. Analytical Chemistry 85: 4651-4657
39. Johansson-Persson A, Barri T, Ulmius M (2013) LC-QTOF/MS metabolomic profiles in human plasma after a 5-week high dietary fiber intake. ANALYTICAL AND BIOANALYTICAL CHEMISTRY 405: 4799-4809.
40. Chernushevich I, Loboda A, Thomson B (2001) An introduction to quadrupole–time-of-flight mass spectrometry. J Mass Spectrometry 36: 849-865.
41. Gomez-Ramos M, Ferrer C, Malato O (2013) Liquid chromatography-high-resolution mass spectrometry for

pesticide residue analysis in fruit and vegetables: Screening and quantitative studies JOURNAL OF CHROMATOGRAPHY A 1287: 24-37.

42. Jupin M, Michiels PJ, Girard FC, Spraul M, Wijmenga SS (2013) NMR identification of endogenous metabolites interacting with fatted and non-fatted human serum albumin in blood plasma: Fatty acids influence the HSA-metabolite interaction. Journal of Magnetic Resonance 228: 81-94.
43. Fukuhara K, Ohno A, Ota Y, Senoo Y, Maekawa K, et al. (2013) NMR-based metabolomics of urine in a mouse model of Alzheimer's disease: identification of oxidative stress biomarkers. Journal of clinical biochemistry and nutrition 52: 133-138.
44. Smolinska A, Blanchet L, Buydens L, Wijmenga S (2012) NMR and pattern recognition methods in metabolomics: From data acquisition to biomarker discovery: A review ANALYTICA CHIMICA ACTA 750: 82-97.
45. Akoka S, Barantin L, Trierweiler M (1999) Concentration measurement by proton NMR using the ERETIC method. Analytical Chemistry 71: 2554-2557.
46. McKay R (2011) How the 1D-NOESY Suppresses Solvent Signal in Metabonomics NMR Spectroscopy: An Examination of the Pulse Sequence Components and Evolution. Concepts in Magnetic Resonance Part A 38A.
47. Del Coco L, Assfalg M, D'Onofrio M (2013) A proton nuclear magnetic resonance-based metabolomic approach in IgA nephropathy urinary profiles. Metabolomics 9: 740-751.
48. Fonville J, Maher A, Coen M (2010) Evaluation of Full-Resolution J-Resolved H-1 NMR Projections of Biofluids for Metabonomics Information Retrieval and Biomarker Identification. Analytical Chemistry 82: 1811-1821.
49. Xi Y, de Ropp J, Jeffrey S, Viant MR (2006) Automated screening for metabolites in complex mixtures using 2D COSYNMR spectroscopy. Metabolomics 2: 221-233.
50. Sandusky P, Appiah-Amponsah E, Raftery D (2011) Use of optimized 1D TOCSY NMR for improved quantitation and metabolomic analysis of biofluids. Journal of Biomolecular NMR 49: 281-290.

51. Koskela H (2009) Quantitative 2D NMR Studies.
52. Mallol R, Rodriguez MA, Heras M, Vinaixa M, Canellas N, et al. (2011) Surface fitting of 2D diffusion-edited H-1 NMR spectroscopy data for the characterisation of human plasma lipoproteins. Metabolomics 7: 572-582.
53. Zhang S, Nagana G, Ye T, Raftery D (2010) Advances in NMR-based biofluid analysis and metabolite profiling. Analyst 135: 1490-1498.
54. Khan M, Ather A (2006) Metabolomics – systematic studies of the metabolic profiling; Elsevier, editor.
55. Kessner D, Chambers M, Burke R, Agusand D, Mallick P (2008) ProteoWizard: open source software for rapid proteomics tools development. Bioinformatics 24: 2534-2536.
56. Smith CA, Want EJ, O'Maille G, Abagyan R, Siuzdak G (2006) XCMS: Processing mass spectrometry data for metabolite profiling using Nonlinear peak alignment, matching, and identification. Analytical Chemistry 78: 779-787.
57. Katajamaa M, Miettinen J, Oresic M (2006) MZmine: toolbox for processing and visualization of mass spectrometry based molecular profile data. Bioinformatics 22: 634-636.
58. Lommen A (2012) Data (pre-)processing of nominal and accurate mass LC-MS or GC-MS data using MetAlign. Methods in molecular biology (Clifton, NJ) 860: 229-253.
59. Kuhl C, Tautenhahn R, Boettcher C, Larson TR, Neumann S (2012) CAMERA: An Integrated Strategy for Compound Spectra Extraction and Annotation of Liquid Chromatography/Mass Spectrometry Data Sets. Analytical Chemistry 84: 283-289.
60. Alonso A, Julia A, Beltran A, Vinaixa M, Diaz M, et al. (2011) AStream: an R package for annotating LC/MS metabolomic data. Bioinformatics 27: 1339-1340.
61. Kopka J, Schauer N, Krueger S, Birkemeyer C, Usadel B, et al. (2005) GMD@CSB.DB: the Golm Metabolome Database. Bioinformatics 21: 1635-1638.
62. Smith CA, O'Maille G, Want EJ, Qin C, Trauger SA, et al. (2005) METLIN - A metabolite mass spectral database. Therapeutic Drug Monitoring 27: 747-751.

63. Wishart DS, Jewison T, Guo AC, Wilson M, Knox C, et al. HMDB 3.0-The Human Metabolome Database in 2013. Nucleic Acids Research 41: D801-D807.
64. Wishart DS (2008) Quantitative metabolomics using NMR. Trends in Analytical Chemistry 27.
65. Skogerson K, Wohlgemuth G, Barupal DK, Fiehn O (2011) The volatile compound BinBase mass spectral database. Bmc Bioinformatics 12.
66. Cuadros-Inostroza A, Caldana C, Redestig H (2009) TargetSearch - a Bioconductor package for the efficient preprocessing of GC-MS metabolite profiling data. Bmc Bioinformatics 10.
67. Jiang W, Qiu Y, Ni Y, Su M, Jia W, et al. (2010) An Automated Data Analysis Pipeline for GC–TOF–MS Metabonomics Studies. Journal of Proteome Research 9: 5974-5981.
68. Kanehisa M, Goto S, Sato Y, Furumichi M, Tanabe M (2012) KEGG for integration and interpretation of large-scale molecular datasets. Nucleic Acids Research 40.
69. Cline M, Smoot M, Cerami E, Kuchinsky A (2007) Integration of biological networks and gene expression data using Cytoscape Nature Protocols 2: 2366 - 2382
70. Matthews L, Gopinath G, Gillespie M, Croft D, de Bono B, et al. (2009) Reactome knowledgebase of human biological pathways and processes. Nucleic Acids Research.
71. Dang L, al e (2009) Cancer-associated IDH1 mutations produce 2-hydroxyglutarate. Nature 462: 739-744.
72. Patti GJ, Yanes O, Shriver LP, Courade J-P, Tautenhahn R, et al. (2012) Metabolomics implicates altered sphingolipids in chronic pain of neuropathic origin. Nature Chemical Biology 8: 232-234.
73. de Caravalho L, Zhao L, Dickinson C, Lima C, Fischer S, et al. (2010) Activity-based metabolomic profiling of enzymatic function: identification of Rv1248c as a mycobacterial 2-hydroxy-3-oxoadipate synthase. Chemical Biology 17: 323-332.
74. García-Romero G, Escobar-Morreale HF (2006) Hyperandrogenism, Insulin Resistance and

Hyperinsulinemia as Cardiovascular Risk Factors in Diabetes Mellitus. Current Diabetes Reviews 2: 39-49.

75. Insenser M, Montes-Nieto R, Murri M, Escobar-Morreale HF (2013) Proteomic and metabolomic approaches to the study of polycystic ovary syndrome. Molecular and Cellular Endocrinology 370: 65-77.
76. Unfer V, Carlomagno G, Dante G, Facchinetti F (2013) Effects of myo-inositol in women with PCOS: a systematic review of randomized controlled trials. Gynecological Endocrinology 28: 509-515.
77. Wright G, Propst A (2012) Polycystic Ovarian Syndrome Management Options. Obstet Gynecol Clin N Am 39: 495-506.
78. Zawadzki J, Dunaif A (1992) Diagnostic criteria for polycystic ovary syndrome: towards a rational approach. Polycystic Ovary Syndrome Boston: Blackwell Scientific Publications;.
79. group. TrEA-sPcw (2004) Revised 2003 consensus on diagnostic criteria and long-term health risks related to polycystic ovary syndrome. Fertility and Sterility 81: 41-47.
80. Goodarzi M (2006) Diagnosis, epidemiology, and genetics of the polycystic ovary syndrome. Best Practice & Research Clinical Endocrinology & Metabolism 20: 193-205.
81. Ibanez L, Lopez-Bermejo A, Diaz M, Victoria Marcos M, de Zegher F (2011) Early Metformin Therapy (Age 8-12 Years) in Girls with Precocious Pubarche to Reduce Hirsutism, Androgen Excess, and Oligomenorrhea in Adolescence. Journal of Clinical Endocrinology & Metabolism 96: E1262-E1267.
82. Murri M, Luque-Ramirez M, Insenser M, Ojeda-Ojeda M, Escobar-Morreale HF (2013) Circulating markers of oxidative stress and polycystic ovary syndrome (PCOS): a systematic review and meta-analysis. Human Reproduction Update 19: 268-288.
83. Asuncion M, Calvo RM, San Millan JL, Sancho J, Avila S, et al. (2000) A prospective study of the prevalence of the polycystic ovary syndrome in unselected Caucasian women from Spain. Journal of Clinical Endocrinology and Metabolism 85: 2434-2438.

84. Escobar-Morreale HF, San Millan JL (2007) Abdominal adiposity and the polycystic ovary syndrome. Trends in Endocrinology and Metabolism 18: 266-272.
85. Azziz R, Carmina E, Dewailly D, Diamanti-Kandarakis E, Escobar-Morreale HF, et al. (2009) The Androgen Excess and PCOS Society criteria for the polycystic ovary syndrome: the complete task force report. Fertil Steril 91: 456-488.
86. Hoffman L, Ehrmann D (2008) Cardiometabolic features of polycystic ovary syndrome. Nature Clinical Practice Endocrinology & Metabolism 4: 215-222.
87. Witchel SF, Recabarren SE, Gonzalez F, Diamanti-Kandarakis E, Cheang KI, et al. (2012) Emerging concepts about prenatal genesis, aberrant metabolism and treatment paradigms in polycystic ovary syndrome. Endocrine 42: 526-534.
88. Salley KE, Wickham EP, Cheang KI, Essah PA, Karjane NW, et al. (2007) Glucose intolerance in polycystic ovary syndrome--a position statement of the Androgen Excess Society. Journal of Clinical Endocrinology and Metabolism 92: 4546-4556.
89. Wild RA, Carmina E, Diamanti-Kandarakis E, Dokras A, Escobar-Morreale HF, et al. (2010) Assessment of cardiovascular risk and prevention of cardiovascular disease in women with the polycystic ovary syndrome: a consensus statement by the Androgen Excess and Polycystic Ovary Syndrome (AE-PCOS) Society. Journal of Clinical Endocrinology and Metabolism 95: 2038-2049.
90. Escobar-Morreale HF, Insenser M, Corton M, San Millán JL, Peral B (2008) Proteomics and genomics: a hypothesis-free approach to the study of the role of visceral adiposity in the pathogenesis of the polycystic ovary syndrome. Proteomics - Clinical Applications 2: 444-456.
91. Corton M, Botella-Carretero JI, Benguria A, Villuendas G, Zaballos A, et al. (2007) Differential gene expression profile in omental adipose tissue in women with polycystic ovary syndrome. J Clin Endocrinol Metab 92: 328-337.
92. Corton M, Botella-Carretero JI, Lopez JA, Camafeita E, San Millan JL, et al. (2008) Proteomic analysis of human omental adipose tissue in the polycystic ovary syndrome

using two-dimensional difference gel electrophoresis and mass spectrometry. Hum Reprod 23: 651-661.

93. Bain JR, Stevens RD, Wenner BR, Ilkayeva O, Muoio DM, et al. (2009) Metabolomics applied to diabetes research: moving from information to knowledge. Diabetes 58: 2429-2443.
94. Escobar-Morreale HF, Sanchon R, San Millan JL (2008) A prospective study of the prevalence of nonclassical congenital adrenal hyperplasia among women presenting with hyperandrogenic symptoms and signs. J Clin Endocrinol Metab 93: 527-533.
95. Azziz R, Carmina E, Dewailly D, Diamanti-Kandarakis E, Escobar-Morreale HF, et al. (2006) Position statement: criteria for defining polycystic ovary syndrome as a predominantly hyperandrogenic syndrome: an Androgen Excess Society guideline. J Clin Endocrinol Metab 91: 4237-4245.
96. Martinez-Garcia MA, Luque-Ramirez M, San-Millan JL, Escobar-Morreale HF (2009) Body iron stores and glucose intolerance in premenopausal women: role of hyperandrogenism, insulin resistance and genomic variants related to inflammation, oxidative stress and iron metabolism. Diabetes Care 32: 1525-1530.
97. Matsuda M, DeFronzo RA (1999) Insulin sensitivity indices obtained from oral glucose tolerance testing: comparison with the euglycemic insulin clamp. Diabetes Care 22: 1462-1470.
98. Palazoglu M, Fiehn O (2009) Metabolite identification in blood plasma using GC/MS and the Agilent Fiehn GC/MS metabolomics RTL library. Agilent Application Note.
99. Dunn WB, Wilson ID, Nicholls AW, Broadhurst D (2012) The importance of experimental design and QC samples in large-scale and MS-driven untargeted metabolomic studies of humans. Bioanalysis 4: 2249-2264.
100. Ek I, Arner P, Bergqvist A, Carlstrom K, Wahrenberg H (1997) Impaired adipocyte lipolysis in nonobese women with the polycystic ovary syndrome: a possible link to insulin resistance? Journal of Clinical Endocrinology and Metabolism 82: 1147-1153.

101. Mozaffarian D, Cao H, King IB, Lemaitre RN, Song X, et al. (2010) Circulating palmitoleic acid and risk of metabolic abnormalities and new-onset diabetes. American Journal of Clinical Nutrition 92: 1350-1358.
102. Setji TL, Holland ND, Sanders LL, Pereira KC, Diehl AM, et al. (2006) Nonalcoholic steatohepatitis and nonalcoholic Fatty liver disease in young women with polycystic ovary syndrome. Journal of Clinical Endocrinology and Metabolism 91: 1741-1747.
103. Newgard CB, An J, Bain JR, Muehlbauer MJ, Stevens RD, et al. (2009) A branched-chain amino acid-related metabolic signature that differentiates obese and lean humans and contributes to insulin resistance. Cell Metab 9: 311-326.
104. Suhre K, Meisinger C, Doring A, Altmaier E, Belcredi P, et al. (2010) Metabolic footprint of diabetes: a multiplatform metabolomics study in an epidemiological setting. PLoS One 5: e13953.
105. Fernstrom JD (2005) Branched-chain amino acids and brain function. Journal of Nutrition 135: 1539S-1546S.
106. Ek I, Arner P, Ryden M, Holm C, Thorne A, et al. (2002) A unique defect in the regulation of visceral fat cell lipolysis in the polycystic ovary syndrome as an early link to insulin resistance. Diabetes 51: 484-492.
107. Faulds G, Ryden M, Ek I, Wahrenberg H, Arner P (2003) Mechanisms behind lipolytic catecholamine resistance of subcutaneous fat cells in the polycystic ovarian syndrome. Journal of Clinical Endocrinology and Metabolism 88: 2269-2273.
108. Layman DK (2003) The role of leucine in weight loss diets and glucose homeostasis. Journal of Nutrition 133: 261S-267S.
109. Altmaier E, Ramsay SL, Graber A, Mewes HW, Weinberger KM, et al. (2008) Bioinformatics analysis of targeted metabolomics--uncovering old and new tales of diabetic mice under medication. Endocrinology 149: 3478-3489.
110. Biolo G, Declan Fleming RY, Wolfe RR (1995) Physiologic hyperinsulinemia stimulates protein synthesis and enhances transport of selected amino acids in human

skeletal muscle. Journal of Clinical Investigation 95: 811-819.

111. Harris SE, Maruthini D, Tang T, Balen AH, Picton HM (2010) Metabolism and karyotype analysis of oocytes from patients with polycystic ovary syndrome. Human Reproduction 25: 2305-2315.
112. Notelovitz M (2002) Androgen effects on bone and muscle. Fertility and Sterility 77 Suppl 4: S34-41.
113. Ciaraldi TP, Aroda V, Mudaliar S, Chang RJ, Henry RR (2009) Polycystic ovary syndrome is associated with tissue-specific differences in insulin resistance. Journal of Clinical Endocrinology and Metabolism 94: 157-163.
114. Ciampelli M, Fulghesu AM, Cucinelli F, Pavone V, Caruso A, et al. (1997) Heterogeneity in beta cell activity, hepatic insulin clearance and peripheral insulin sensitivity in women with polycystic ovary syndrome. Human Reproduction 12: 1897-1901.
115. Fenkci V, Fenkci S, Yilmazer M, Serteser M (2003) Decreased total antioxidant status and increased oxidative stress in women with polycystic ovary syndrome may contribute to the risk of cardiovascular disease. Fertil Steril 80: 123-127.
116. Samino S, Revuelta-Cervantes J, Vinaixa M, Rodriguez MA, Valverde AM, et al. (2013) A 1H NMR metabolic profiling to the assessment of protein tyrosine phosphatase 1B role in liver regeneration after partial hepatectomy. Biochimie 95: 808-816.
117. Deja S, Barg E, Mlynarz P, Basiak A, Willak-Janc E (2013) (1)H NMR-based metabolomics studies of urine reveal differences between type 1 diabetic patients with high and low HbAc1 values. Journal of pharmaceutical and biomedical analysis 83: 43-48.
118. van Vliet E, Eixarch E, Illa M, Arbat-Plana A, Gonzalez-Tendero A, et al. (2013) Metabolomics Reveals Metabolic Alterations by Intrauterine Growth Restriction in the Fetal Rabbit Brain. Plos One 8.
119. Ruiz-Aracama A, Peijnenburg A, Kleinjans J, Jennen D, van Delft J, et al. (2011) An untargeted multi-technique metabolomics approach to studying intracellular

metabolites of HepG2 cells exposed to 2,3,7,8-tetrachlorodibenzo-p-dioxin. Bmc Genomics 12.

120. Shaham O, Slate NG, Goldberger O, Xu Q, Ramanathan A, et al. (2010) A plasma signature of human mitochondrial disease revealed through metabolic profiling of spent media from cultured muscle cells. Proceedings of the National Academy of Sciences of the United States of America 107: 1571-1575.
121. Koek M, Jellema R, Van der Greef J, Tas A, Hankemeier T (2011) Quantitative metabolomics based on gas chromatography mass spectrometry: status and perspectives. Metabolomics 7: 307-328.
122. Yanes O, Clark J, Wong DM, Patti GJ, Sanchez-Ruiz A, et al. Metabolic oxidation regulates embryonic stem cell differentiation. Nature Chemical Biology 6: 411-417.
123. Gonzalez F, Nair S, Daniels J, Basal E, Schimke J, et al. (2012) Hyperandrogenism sensitizes leuocytes to hyperglycemia to promote oxidative stress in lean reproductve-age women Journal of Clinical Endocrinology & Metabolism 97: 2836-2843.
124. Gonzalez F, Rote N, Minium J, Kirwan J (2006) Reactive oxygen species-induced oxidative stress in the development of insulin resistnace and hyperandrogenism in polycystic ovary syndrome. Journal of Clinical Endocrinology & Metabolism 91: 336-340.

CHAPTER 6. ANNEX

Biochimie 95 (2013) 808–816

Contents lists available at SciVerse ScienceDirect

Biochimie

journal homepage: www.elsevier.com/locate/biochi

Research paper

A ^{1}H NMR metabolic profiling to the assessment of protein tyrosine phosphatase 1B role in liver regeneration after partial hepatectomy

Sara Samino [a,b,1], Jesús Revuelta-Cervantes [b,c,1], Maria Vinaixa [a,b], Miguel Ángel Rodríguez [a,b,*], Ángela M. Valverde [b,c], Xavier Correig [a,b]

[a] Metabolomics Platform, IISPV, Universitat Rovira i Virgili, Avda. Països Catalans, 26, 43007 Tarragona, Spain
[b] Centro de Investigación Biomédica en Red de Diabetes y Enfermedades Metabólicas Asociadas (CIBERDEM), Barcelona, Spain
[c] Instituto de Investigaciones Biomédicas Alberto Sols, CSIC-UAM, Madrid, C/Arturo Duperier, 4, 28029 Madrid, Spain

ARTICLE INFO

Article history:
Received 30 May 2012
Accepted 28 November 2012
Available online 12 December 2012

Keywords:
PTP1B
Diabetes
Liver regeneration
^{1}H NMR spectroscopy
Metabolomic

ABSTRACT

Protein tyrosine phosphatase 1B (PTP1B) is a negative regulator of the tyrosine kinase growth factor signaling pathway, which is involved in major physiological mechanisms such as liver regeneration. We investigate early hepatic metabolic events produced by partial hepatectomy (PHx) for PTP1B deficient (PTP1B KO) and wild type (WT) mice using proton nuclear magnetic resonance spectroscopy. Metabolic response of the two genotypes produced 24 h upon PHx is compared using magic angle spinning high-resolution nuclear magnetic resonance (^{1}H-HR-MAS-NMR) on intact liver tissues. In addition, genotype-associated metabolic profile changes were monitored during the first 48 h after PHx using high-resolution nuclear magnetic resonance (^{1}H-HR-NMR) on liver extracts. A marked increase of lipid-related signals in regenerating livers was observed after 24 h PHx in either intact tissues or liver extracts studies. In spite of this common initial metabolic response, results obtained 48 h after PHx on liver extracts indicate a genotype-differential metabolic pattern. This metabolic pattern resulted in line with well known regenerative features such as more sustained cell proliferation, a better management of lipids as energy fuel and lessened liver injury for PTP1B KO mice as compared to WT. Taken together, these findings suggest the metabolic basis to the pivotal role of PTP1B in liver regeneration.

1. Introduction

The role of Protein Tyrosine Phosphatase 1B (PTP1B) as a regulator of signal transduction by dephosphorylation of Receptor Tyrosine Kinases (RTK) is well known [1–5] and it has been established as a regulator of many signaling pathways associated with human diseases, including diabetes [6,7] obesity [8,9] and cancer [10,11]. Different studies have explored regulatory links between PTP1B and the insulin receptor [12–14]. The development of PTP1B knockout mice [15,16] has allowed further study of the role of PTP1B in diabetes and obesity [17]. PTP1B KO mice have been shown to grow up normally without any significant difference in weight gain or fertility as compared to wild type (WT) mice [15]. Furthermore, PTP1B KO mice maintain lower glucose concentrations with significantly reduced amounts of insulin, displaying a tissue-specific increased sensitivity to insulin [15,16]. Currently, the inhibition of PTP1B is under study as a potential therapy target for diabetes [2,3,18–20]. In addition, PTP1B has been implicated in the control of cell adhesion, which regulates interactions with the extracellular matrix at focal adhesion complexes [21–23].

On the other hand, since liver regeneration involves tyrosine phosphorylation-mediated signaling we have recently investigated the role of PTP1B in hepatic regeneration by performing partial hepatectomy (PHx) in WT and PTP1B KO mice [24]. The essential role of PTP1B in the response to the cytokines and growth factors that initiate hepatic regeneration was assessed, as well as in the control of its termination. PTP1B KO mice triggered an earlier response to hepatic injury as compared to WT, i.e., a more rapid increase in intrahepatic lipid accumulation; an accelerated phosphorylation of JNK1/2 and STAT3; enhanced EGFR and HGFR mediated signaling and enhanced expression of proliferative cell markers such as the proliferative cell nuclear antigen (PCNA) and cyclins D1 and E. Based on this evidence, a metabolomic study was planned to investigate the metabolic rearrangements produced in

* Corresponding author. Metabolomics Platform (URV-CIBERDEM), Campus Sescelades, SRCIT Universitat Rovira i Virgili, Avda. Països Catalans, 26, 43007 Tarragona, Spain. Tel.: +34 977 256 587; fax: +34 977 55 8261.
E-mail address: miguelangel.rodriguez@urv.cat (M.Á. Rodríguez).
URL: http://www.metabolomicsplatform.com
[1] These authors contributed equally to this work.

0300-9084/$ – see front matter © 2012 Elsevier Masson SAS. All rights reserved.
http://dx.doi.org/10.1016/j.biochi.2012.11.015

Annex

the liver upon PHx in PTP1B KO with the rationale that they would further shed light on the metabolic processes involved in such early triggered response. Thus, a time-course study along the initial stages of liver regeneration using an NMR-based untargeted metabolomic approach was assayed on partial hepatectomized PTP1B KO and WT mice. The first part of the study was performed on intact liver tissue using High Resolution Magic Angle Spinning (HR-MAS) NMR spectroscopy in which PTP1B KO was compared to WT mice at time points 0 and 24 h after PHx. The second part of the study consisted of the same comparison involving an extra time point (48 h after injury), performed on aqueous (acetonitrile/water) and lipidic (chloroform/methanol) liver extracts.

2. Materials and methods

2.1. Animals

The animals were 12-week-old male wild type (WT, *Ptpn1*$^{+/+}$) and PTP1B Knock Out (PTP1B KO, *Ptpn1*$^{-/-}$) mice of the mixed genetic background 129 sv x C57/BL6 as previously described [24]. Animal care followed the recommendations of the Federation of European Laboratory Animal Science Associations (FELASA) in terms of health monitoring. The use of animals in these experimental procedures was approved by the Consejo Superior de Investigaciones Cientificas (CSIC) Animal Care and Use Committee. Mice were fed a standard chow diet.

2.2. Partial hepatectomy surgery procedure

A subset of four to seven animals for each condition was subjected to a standard 70% PHx under isoflurane anesthesia, as previously described [25]. Analgesia was provided prior to surgery by subcutaneous injection of buprenorphine 50 μg/kg (Buprex, Schering-Plough). Briefly, the liver lobes were extruded, the median and left lateral lobes were ligated and removed as described above, and the wounds were sutured. The left lateral lobe was immediately frozen at −80 °C and stored until assayed. Animals were provided with water and food *ad libitum* and kept in a warm environment to avoid hypothermia. The remaining lobes were collected 24 or 48 h after PHx. Samples of tissue consisting of the distal portion of the anterior right lobe were weighed, chopped into smaller pieces of the desired size and immediately frozen and stored at −80 °C.

2.3. Oil red O staining

Portions of regenerating liver were fixed, embedded in O.C.T. (Tissue Tek, Sakura Finetek Europe) and flash frozen in liquid nitrogen. 5 μm tissue slices were stained with oil red O and counterstained with hematoxylin to reveal the presence of triglycerides and lipids in the liver.

2.4. Sample preparation and HR-MAS (ex-vivo) intact liver tissue measurements

Individual frozen tissues samples (9.8–19.5 mg) were placed in an insert for a 4 mm outer diameter ZrO_2 rotor, limiting the rotor inner volume to 20 μl. Then, the insert was filled with cooled D_2O, sealed and subsequently inserted into the ZrO_2 rotor. HR-MAS spectra were recorded on a Bruker Avance III 500 spectrometer operating at a proton frequency of 500.13 MHz. The instrument was equipped with a 4 mm triple resonance (^{1}H, ^{13}C, ^{31}P) gradient HR-MAS probe. A Bruker Cooling Unit (BCU-Xtreme) was used to keep the sample temperature at 4 °C. Samples properly prepared with D_2O and shimmed were spun at 4 kHz to keep the rotation sidebands out of the spectral region of interest [26]. One-dimensional (1D) ^{1}H spectra were acquired using a noesy presat sequence for water suppression. A total of 128 scans with a 90° flip angle were accumulated in 23 min, with a 7000 Hz (14 ppm) of spectral width and 8.0 s of relaxation delay to ensure full relaxation of magnetization between scans.

2.5. Liver tissue extracts preparation and high-resolution liquid NMR (in-vitro) measurement

Around 100 mg of liver (left lateral or anterior right lobes for baseline or experimental time-point, respectively) were homogenized in 2 ml of a cold mixture (CH_3CN:H_2O 1:1 v/v T = 0 °C, 5 min) by using a T10 basic Ultra-Turrax Disperser (IKA, Staufen, Germany). The resulting suspension was centrifuged (5000 × g, 15 min, T = 4 °C) and the supernatant was carefully transferred to a new eppendorf tube. This step was repeated three times and the combined aqueous phases were frozen, lyophilized and stored at −80 °C until further analysis. The resulting pellet was dried and extracted with 1 ml of a chloroform:methanol mixture at room temperature for 20 min (CH_3Cl:MeOH 3:1 v/v). Then, it was centrifuged for 5 min at 6000 × g at 4 °C and the supernatant (lipidic phase) was collected in a new eppendorf tube. The lipidic phase was dried under a nitrogen stream and stored at −80 °C. For NMR measurements, the hydrophilic extracts were reconstituted in 600 μl of D_2O containing 0.58 mM trisilylpropionic acid (TSP). The lipophilic extracts were subsequently extracted in 700 μl of a solution $CDCl_3$/CD_3OD (2:1) containing 0.73 mM tetramethylsilane (TMS) and samples were then vortexed, homogenized for 20 min, centrifuged for 15 min at 6000 × g at room temperature and transferred into 5 mm NMR tubes.

One- and two-dimensional ^{1}H NMR spectra were measured at a 600.20 MHz frequency using an Avance III-600 Bruker spectrometer equipped with an inverse TCI 5 mm cryoprobe®. For the 1D aqueous extract spectra, one-dimensional (1D) Nuclear Overhauser Effect Spectroscopy with a spoil gradient (noesygppr1d) was used. Solvent presaturation with low irradiation power (10 Hz) was applied during recycling delay and mixing time (tm = 100 ms) to suppress residual water. A total of 256 transients were collected across 12 kHz spectral width at 300 K into 64 k data points, and exponential line broadening of 0.3 Hz was applied before Fourier transformation. A recycling delay time of 8 s was applied between scans to ensure correct quantification. In the case of lipophilic extracts, a 90° pulse with a presaturation sequence (zgpr) was used. We performed measurements at 287 K, shifting the residual water signal to 4.65 ppm to allow for the quantification of the characteristic glycerol-backbone signals. In addition, residual water was pre-saturated during recycling delay (RD = 8 s) using a low irradiation power (10 Hz). A total of 256 FIDs of 12 kHz of spectral width were collected into 64 k data points, and exponential line broadening of 0.3 Hz was applied before Fourier transformation. The frequency spectra were phased, baseline corrected and then calibrated (TMS or TSP, 0.0 ppm) using TopSpin software (version 2.1, Bruker®, Germany).

2.6. Metabolite identification and quantification

Resonance assignments were done on the basis of literature values and different database search engines (Bioref AMIX 3.8 database from Bruker®; Chenomx NMRSuite 7.5 from Chenomx Inc. and Human Metabolome Database, HMDB http://www.hmdb.ca). Structural confirmations were based on two-dimensional (2D) ^{1}H–^{1}H COrrelation SpectroscopY (COSY) and 2D ^{1}H, ^{13}C Heteronuclear Single Quantum Correlation (HSQC). We selected 1D NMR regions based on our previously published 1D ^{1}H NMR lipidic and aqueous

resonances assignments and their structural correlates [27]. The area under the regions identified within the HR-MAS spectra was integrated using AMIX 3.8 software package (Bruker GmbH, Germany). Each area was then normalized using the Electronic REference To access In vivo Concentrations (ERETIC) signal [28] and weight of tissue in each sample. In lipidic and aqueous extracts, selected peaks in the 1D-NMR spectra were also integrated using the AMIX 3.8 software package. The concentration of each metabolite in either water-soluble extracts or lipid-soluble extracts was assessed according to the methodology previously described [29]. Finally, the fatty acyl chain composition in terms of molar percentage was calculated as previously described [27].

2.7. Data processing and statistical analysis

In the case of HR-MAS measurements integrated regions were used as the input matrix for multivariate data analysis. In the case of liver extracts, we used the calculated concentrations of metabolites identified in both lipophilic and hydrophilic extracts arranged together in one single data matrix. In both cases, a first overview of the global metabolic pattern response was explored using Principal Component Analysis (PCA) on data previously scaled to unit variance. Additionally, in order to study the genotype differential metabolic response to PHx, a two-factor ANOVA on rank-transformed data was calculated for each metabolite identified on the extracts ($p < 0.05$ was considered for statistical significance). Data (pre-) processing, data analysis and statistical calculations were performed with Matlab (Matlab version 6.5.1, Release 13, The Mathworks, 2003 and the PLS Toolbox, version 4.2).

3. Results and discussion

3.1. Histological and hepatosomal index analysis

Visual inspection of red on oil red O staining images (Fig. 1) showed similar hepatic fat content in liver sections of both mouse genotypes before surgery. Conversely, there appeared to be more oil red O staining in the KO mice at 24 and 48 h after PHx, indicating an increase in hepatic fat content in PTP1B KO animals as compared to their WT counterparts (Fig. 1).

The recovery of the liver mass following PHx was monitored through calculation of the hepatosomal index (the percentage ratio of liver to body weight) at sacrifice (Fig. 2). In the case of PTP1B KO animals, the hepatosomal index increased after 24 h PHx as compared to WT strain indicating accelerated liver regeneration. However, 48 h after PHx, hepatomal index no longer differs significantly between WT and KO mice.

3.2. Preliminary HR-MAS study in intact liver tissue

At 24 h after PHx, the regenerating livers of both PTP1B KO and WT mice primarily showed a marked increase of signals related to

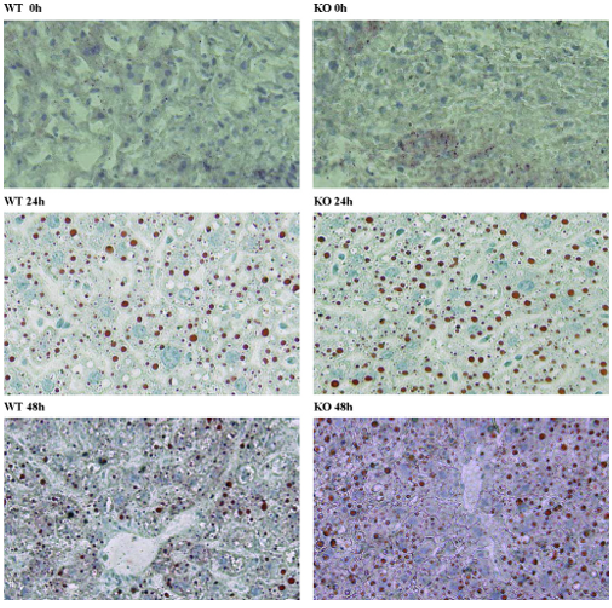


Fig. 1. Lipid accumulation in the liver due to regeneration. Oil red O staining of liver sections from WT and PTP1B KO mice at indicated time (hours) after PHx. Original magnification ×10.

Annex

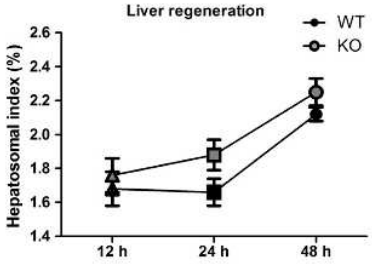


Fig. 2. Liver re-growth in hepatectomized animals. Hepatosomal index (expressed as percentage of liver weight/body weight ratio) through regenerative process is shown at indicated hours after surgery. Values are expressed as mean ± sem.

lipid resonance such as polyunsaturated fatty acid (PUFA) signals arising at δ 2.5–2.8 ppm, allylic protons at 1.9–2.1 ppm, fatty acid α protons at δ 1.5–1.7 ppm, methylene protons at δ 1.3–1.4 ppm and ω-CH_3 at 0.8 ppm (Fig. 3). Notably, glycerol signals of triglycerides backbone arising at δ 4–4.3 ppm were more prominent in WT animals as compared to PTP1B KO animals.

The two genotypes livers at baseline (0 h, before PHx) and at 24 h after PHx were compared using PCA calculated on their corresponding HR-MAS spectra. Main variation along Principal Component 1 (PC1) explaining 59.25% of the variance, segregates samples according to time points after PHx (Fig. 4A) and no differences between the two strains can be inferred. To assess metabolites implicated in the variations observed after PHx, the PC1 loadings plot was studied (Fig. 4B). Spectral regions accounting for higher absolute values in the bar loadings plot exert higher influence in the PCA model. Thus, regenerating livers at 24 h after surgery presented a higher contribution of fatty acyl-related signals, triglycerides and cholesterol. Conversely, they presented depleted areas of regions identified as glucose-, glycerol- and phospholipid-related signals.

It should be considered that one-dimensional HR-MAS of intact tissue has inherent assignment and quantification challenges due to the high overlap of small molecule signals with many broad signals derived from macromolecules and lipids. This holds particularly true in this study where broad lipid signals were highly prominent because of lipid accumulation within the first 24 h after PHx. To account for overlapping signals and to determine the detailed lipid composition, we performed the same comparisons on extracted liver tissues.

3.3. Global metabolic rearrangements in PHx mice as assessed in liver extracts

Liver extracts analysis involved an extra-time point. Hence, the two genotypes livers at baseline, at 24 and at 48 h after PHx were compared using PCA calculated on the matrix resulting from metabolites identified both in lipid and water-soluble extracts. Fig. 5A shows the 1D-PC2 trajectory score plot depicting time-dependent metabolic status after PHx. At 24 h time-point (squares) both genotypes showed an increased PC2 score value respect to baseline (triangles), being this increase more prominent for WT mice. This results in good agreement with our HR-MAS analysis. From 24 to 48 h after PHx, PTP1B deficient mice hold the same PC2 values (black circles). Conversely, this trajectory was not the same for WT genotype after 48 h surgery (gray circles) showing a PC2 scores value similar to baseline levels. The PC2 loadings bar plot (Fig. 5B) showed that livers at 24 h after PHx are characterized by higher levels of oleic acid and 3-hydroxybutyric acid, among other metabolites. Conversely, livers at baseline show higher levels of glucose, glycerol, and PUFAs including arachidonic acid (ARA), eicosapentaenoic acid (EPA) or docosahexaenoic acid (DHA). In order to gain insight into metabolic rearrangements differently occurred between WT and PTP1B KO mice that underwent PHx, we studied metabolites levels in more detail.

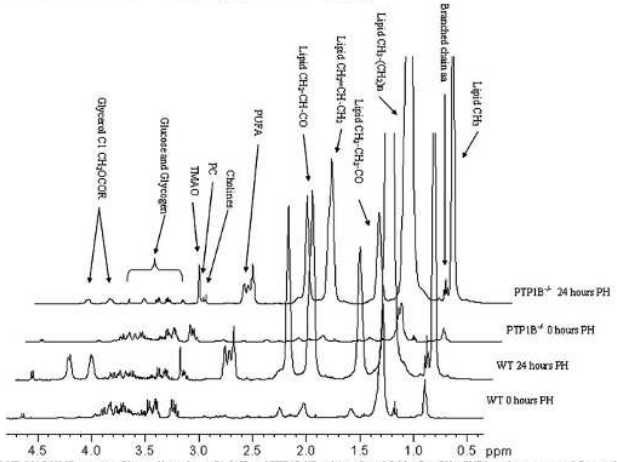


Fig. 3. Representative ^{1}H-HR-MAS NMR spectra of intact liver tissue for WT and PTP1B KO mice at 0 and 24 h after PHx. PUFA: polyunsaturated fatty acids, PC: phosphatidylcholine, TMAO: trimethylamine-N-oxide.

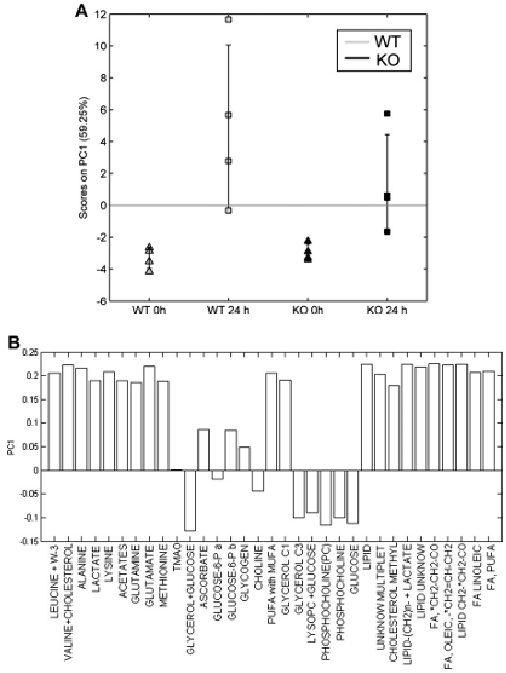


Fig. 4. (A) PC1 score plot of ^{1}H-HR-MAS NMR spectral data for WT (gray spots) and KO (black spots) at time points 0 h (△, WT $n = 4$ and KO = 4) and 24 h (□, WT $n = 4$ and KO = 4) after PHx. The area of 32 spectral regions (variables) was considered as input matrix for PCA. Error bar indicates PC1 sem error. Mice are segregated along PC1 according to regeneration time point. (B) Loadings bar plot showing the NMR regions responsible for such trends.

3.4. Differential metabolic profiles in PTP1B KO vs. WT hepatectomized mice

3.4.1. Statistical analysis

The first, early metabolic events occurring in response to PHx were studied comparing WT and PTP1B KO genotypes in relation to their metabolic response between the baseline and the first 24 h after PHx. The 0–24 h changes observed in ARA + EPA, w-3 fatty acids (DHA + EPA + linoleic), glutamate and methionine differed significantly between the two genotypes. Between 24 and 48 h, the changes in a larger number of metabolites were different between genotypes, most of them related to the saturation and composition of fatty acyl chains (Table 1, p-interaction < 0.05, two-factor ANOVA).

Additionally, the actual concentrations of metabolites significantly varied from baseline to 24 h and from 24 to 48 h after PHx in the independent WT group (Supporting Information, Table S1); and concentrations in the independent PTP1B deficient animals (Supporting Information, Table S2) are also provided.

3.4.2. Lipidic metabolites

During the first 24 h after PHx, levels of both mono- or di-unsaturated fatty acids such as oleic and linoleic were found to be significantly increased in the two genotypes (t-test, $p < 0.05$, Tables S1 and S2). However, from 24 to 48 h after PHx, the two genotypes behave differently respective to these two acids (Table 1): whilst in KO mice, oleic and linoleic acids levels were sustained from 24 to 48 h after PHx, WT mice showed a significant depletion (t-test, $p < 0.05$, Table S1) for these two acids returning to baseline levels (Fig. 6A, B). Because increased levels of oleic acid are indicative of cell proliferation [30,31], these results indicate that liver regeneration process is prolonged 48 h after PHx in PTP1B KO but not in WT animals. This results in good agreement with previously described findings [24].

Annex

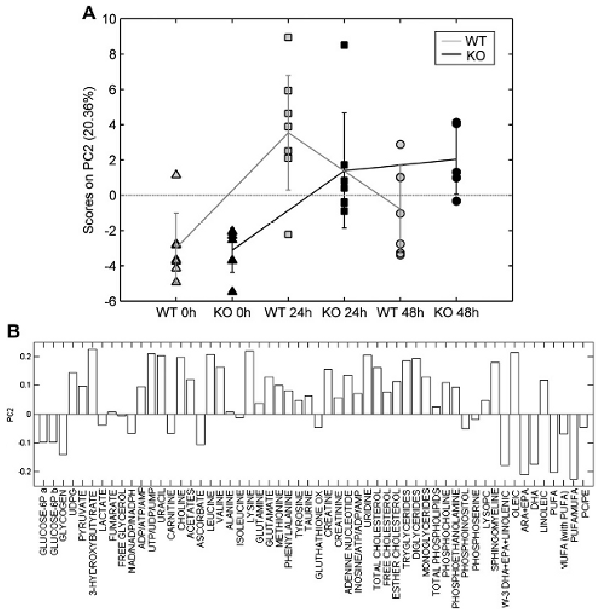


Fig. 5. (A) PC2 score plot based on the metabolic patterns obtained from integrated NMR spectral regions of aqueous and lipophilic liver tissue extracts (for a total of 56 variables). The two genotypes (WT in gray and PTP1B KO in black) are represented at time points 0 h (△, WT $n = 7$ and KO $n = 7$), 24 h (□, WT $n = 8$ and KO $n = 7$) and 48 h (◇, WT $n = 7$ and KO $n = 5$) after PHx. Group centroids are indicated, together with lines representing the genotype differential metabolic trajectories in the reduced scores space after PHx. (B) PC2 loadings plot revealing the pattern of metabolites implicated in the PHx induced changes.

After a common initial significant depletion of DHA, ARA + EPA and ω-3 levels at the 24 h time point (Fig. 6C–E) (*t*-test, $p < 0.05$, Tables S1 and S2), levels of PUFA were restored 48 h after PHx in WT animals. Conversely, they remained depleted for the *Ptpn1*$^{-/-}$. It is worth to mention that, from 24 to 48 h, changes in PUFA/MUFA levels were significantly different between the two genotypes (Table 1, Fig. 6F). Therefore, the inhibitory effect of PUFA on cell proliferation [30,32] further supports the observed sustained liver regeneration in PTP1B KO as compared to WT mice.

The role of triglycerides in hepatic regeneration has long been recognized [33–35]. In addition, different studies have demonstrated an activation of adipogenic genes prior to liver regeneration [35,36]. After PHx, the peripheral adipose tissues release fatty acids to enhance hepatic fatty acid uptake, which in

Table 1
Metabolites that varied significantly according to genotype from 0 to 24 h and from 24 to 48 h after PHx (p-interaction < 0.05, two-factors ANOVA). ARA: arachidonic acid, EPA: eicosapentaenoic acid, DHA: docosahexaenoic acid, PUFA: polyunsaturated fatty acid, MUFA: monounsaturated fatty acid.

0–24 h after PHx				24–48 h after PHx			
Metabolite	*p*-Value	Fold change WT	Fold change KO	Metabolite	*p*-Value	Fold change WT	Fold change KO
Glutamate	3.8E−02	2.33	1.39	3-Hydroxybutyric acid	4.8E−02	−2.05	1.43
Methionine	1.4E−02	3.21	1.25	Uracil	3.2E−03	−3.57	1.84
ω-3 (DHA + EPA + Linolenic)	4.6E−02	−1.85	−1.27	Triglycerides	6.5E−04	−7.95	1.16
ARA + EPA	3.8E−02	−3.59	−2.48	Diglycerides	3.6E−02	−3.42	−2.10
				ω-3 (DHA + EPA + Linolenic)	1.0E−02	1.87	−1.05
				Oleic	2.3E−03	−1.60	1.16
				ARA + EPA	4.6E−03	2.52	−1.09
				DHA	3.7E−02	2.19	1.13
				Linoleic	1.4E−02	−1.31	−1.07
				PUFA	3.8E−03	2.23	1.04
				PUFA/MUFA	1.2E−03	2.04	−1.02

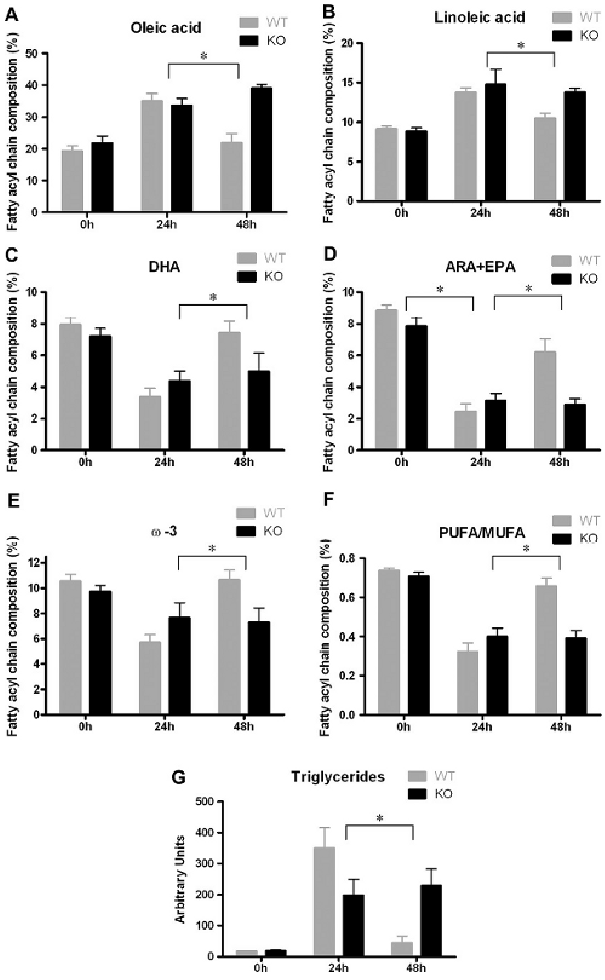


Fig. 6. Evolution of lipophilic liver extract metabolites showing a significantly different behavior with PHx in the two genotypes. *(p-interaction < 0.05, two-factors ANOVA).

turn promotes hepatic lipogenesis and leads to a rapid accumulation of intracellular triglycerides within the regenerating liver [34]. Our results show that, after an initial significant rise of hepatic triglycerides peaking at 24 h post-PHx in both genotypes, WT levels of triglycerides were returned to normal levels after 48 h post PHx (t-test, $p < 0.05$, Table S1), while high levels were prolonged in PTP1B KO animals at the same time point (Fig. 6G).

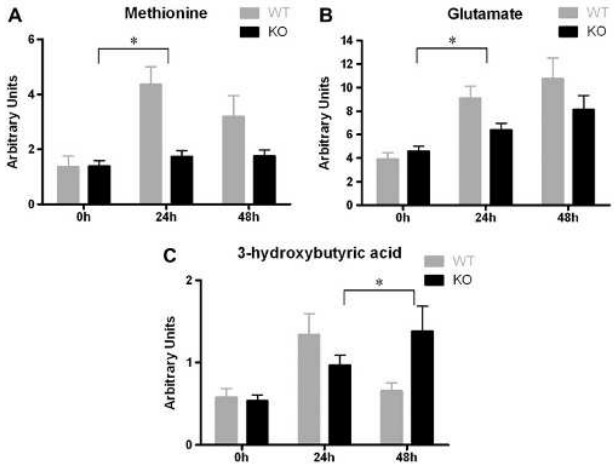


Fig. 7. Evolution of water-soluble liver extract metabolites showing a significantly different behavior with PHx in the two genotypes. *(p-interaction < 0.05, two-factor ANOVA).

3.4.3. Aqueous metabolites

In WT animals, methionine levels were significantly increased 24 h after PHx with respect to the baseline (t-test, $p < 0.05$, Table S1), while there was no variation in PTP1B KO mice. Thus, methionine levels were affected differently according to the genotype (Table 1, Fig. 7A). Methionine is a substrate for S-adenosyl methionine transferase, which converts methionine into S-adenosyl methionine (SAM). In rat livers after a PHx, SAM levels are dramatically reduced shortly afterward, coincident with the onset of DNA synthesis and the induction of early response genes [37,38]. In addition, this transient fall in hepatic SAM levels is known to precede regeneration after PHx and is considered to be necessary for this process to occur [39,40]. Due to the pivotal role of SAM in liver regeneration and due to the fact that methionine accounts for SAM availability, we assessed the variation in methionine at one extra-point: 12 h after PHx. Changes in methionine levels up to 48 after PHx are shown in Figure S1. Increased levels of methionine, peaking at 12 h after PHx for both genotypes, showed an earlier fall in KO as compared to WT animals, which is suggestive of an earlier regenerative metabolic response in these animals as compared to WT animals [40].

Transaminases play an important role and are indicative of liver injury [41]. Glutamate, a byproduct of transamination activity, is increased in both strains 24 h after PHx (Tables S1 and S2 and Fig. 7B). Nevertheless, PTP1B deficiency diminished the increase in glutamate, suggesting lower transamination activity. These results are consistent with our previous study [24], in which transaminases are measured in liver, and levels of transaminases are decreased more rapidly along the regeneration process in PTP1B KO mice.

The significant increase in 3-hydroxybutyric acid at the 24 h time point for both strains, together with depleted glucose and glycogen levels (t-test, $p < 0.05$, Tables S1 and S2), is indicative that both strains are using lipids as an energy fuel (Fig. 7C).

However, while at 48 h after PHx, 3-hydroxybutyric levels are decreased in WT mice, they are increased in PTP1B KO mice. These results suggest that the usage of lipids as an energy fuel lasts for a longer time period in absence of PTP1B.

4. Conclusions

A dramatic increase of intrahepatic lipids in mice has been demonstrated 24 h after PHx using ^{1}H HR-MAS-NMR on intact tissue. In addition, studies performed on liver extracts demonstrated that, even though early metabolic changes from baseline to 24 h after surgery point in the same direction in both genotypes, these changes were less pronounced in the case of PTP1B KO animals. However from 24 to 48 h after PHx, a genotype-related differential metabolic response was determined. Conversely to WT, PTP1B KO mice were shown to maintain oleic levels while they did not increase PUFA levels 48 h after PHx. In agreement with the previously reported prolongation of genetic and molecular response elicited by PHx [24], this result suggests of a more sustained cell proliferation. Moreover, our metabolic parameters point to a lessened extent of liver damage during the regeneration process and improved energetic effectiveness. Taken together, these findings establish a pivotal role for PTP1B in the metabolic response that occurs along the early stages in hepatic regeneration.

This work provides a proof-of-concept as to whether the two NMR metabolic profiling approaches used have proven to be successful in liver phenotyping of experimental animal models. Both ^{1}H-HR-MAS NMR on intact liver tissues and ^{1}H NMR liver extract profiling have been shown to provide a biochemical snapshot of the metabolic changes that occur after PHx, allowing for a global assessment of the metabolic status of the liver.

Acknowledgements

The authors would like to thank Dr. Catherine Heeney for English revision. This work was supported by grants SAF2009-08114 from Ministerio de Ciencia e Innovación (MICINN, Spain) to

A.M.V. and by CIBER de Diabetes y Enfermedades Metabólicas Asociadas (CIBERDEM). CIBER de Diabetes y Enfermedades Metabólicas Asociadas is an initiative of Instituto de Salud Carlos III, MICINN, Spain. Sara Samino is the recipient of a fellowship from Institut d'Investigació Sanitària Pere Virgili (IISPV).

Appendix A. Supplementary data

Supplementary data related to this article can be found at http://dx.doi.org/10.1016/j.biochi.2012.11.015.

References

[1] J.N. Andersen, N.K. Tonks, Protein tyrosine phosphatase-based therapeutics: lessons from PTP1B, Protein Phosphatases 5 (2004) 201–230.
[2] A.P. Combs, Recent advances in the discovery of competitive protein tyrosine phosphatase 1B inhibitors for the treatment of diabetes, obesity, and cancer, Journal of Medicinal Chemistry 53 (2010) 2333–2344.
[3] B.L. Seely, P.A. Staubs, D.R. Reichart, P. Berhanu, K.L. Milarski, A.R. Saltiel, J. Kusari, J.M. Olefsky, Protein tyrosine phosphatase 1B interacts with the activated insulin receptor, Diabetes 45 (1996) 1379–1385.
[4] M. Stuible, J.V. Abella, M. Feldhammer, M. Nossov, V. Sangwan, B. Blagoev, M. Park, M.L. Tremblay, PTP1B targets the endosomal sorting machinery: dephosphorylation of regulatory sites on the endosomal sorting complex required for transport component STAM2, Journal of Biological Chemistry 285 (2010) 23899–23907.
[5] M. Stuible, M.L. Tremblay, In control at the ER: PTP1B and the down-regulation of RTKs by dephosphorylation and endocytosis, Trends in Cell Biology 20 (2010) 672–679.
[6] F. Ahmad, P.M. Li, J. Meyerovitch, B.J. Goldstein, Osmotic loading of neutralizing antibodies demonstrates a role for protein-tyrosine phosphatase 1B in negative regulation of the insulin action pathway, Journal of Biological Chemistry 270 (1995) 20503–20508.
[7] B.J. Goldstein, F. Ahmad, W. Ding, P.M. Li, W.R. Zhang, Regulation of the insulin signalling pathway by cellular protein-tyrosine phosphatases, Molecular and Cellular Biochemistry 182 (1998) 91–99.
[8] J.M. Zabolotny, K.K. Bence-Hanulec, A. Stricker-Krongrad, F. Haj, Y. Wang, Y. Minokoshi, Y.B. Kim, J.K. Elmquist, L.A. Tartaglia, B.B. Kahn, B.G. Neel, PTP1B regulates leptin signal transduction in vivo, Developmental Cell 2 (2002) 489–495.
[9] S. Zhang, Z.Y. Zhang, PTP1B as a drug target: recent developments in PTP1B inhibitor discovery, Drug Discovery Today 12 (2007) 373–381.
[10] T. Fukada, N.K. Tonks, The reciprocal role of Egr-1 and Sp family proteins in regulation of the PTP1B promoter in response to the p210 Bcr-Abl oncoprotein-tyrosine kinase, Journal of Biological Chemistry 276 (2001) 25512–25519.
[11] K.R. LaMontagne, G. Hannon, N.K. Tonks, Protein tyrosine phosphatase PTP1B suppresses p210 bcr-abl-induced transformation of Rat-1 fibroblasts and promotes differentiation of K562 cells, Proceedings of the National Academy of Sciences of the United States of America 95 (1998) 14094–14099.
[12] F. Escriva, A. Gonzalez-Rodriguez, E. Fernandez-Millan, C.M. Rondinone, C. Alvarez, A.M. Valverde, PTP1B deficiency enhances liver growth during suckling by increasing the expression of insulin-like growth factor-I, Journal of Cellular Physiology 225 (2010) 214–222.
[13] H. Maegawa, R. Ide, M. Hasegawa, S. Ugi, K. Egawa, M. Iwanishi, R. Kikkawa, Y. Shigeta, A. Kashiwagi, Thiazolidine derivatives ameliorate high glucose-induced insulin-resistance via the normalization of protein-tyrosine-phosphatase activities, Journal of Biological Chemistry 270 (1995) 7724–7730.
[14] A.S. Mueller, A.C. Bosse, E. Most, S.D. Klomann, S. Schneider, J. Pallauf, Regulation of the insulin antagonistic protein tyrosine phosphatase 1B by dietary Se studied in growing rats, Journal of Nutritional Biochemistry 20 (2009) 235–247.
[15] M. Elchebly, P. Payette, E. Michaliszyn, W. Cromlish, S. Collins, A.L. Loy, D. Normandin, A. Cheng, J. Himms-Hagen, C.C. Chan, C. Ramachandran, M.J. Gresser, M.L. Tremblay, B.P. Kennedy, Increased insulin sensitivity and obesity resistance in mice lacking the protein tyrosine phosphatase-1B gene, Science 283 (1999) 1544–1548.
[16] L.D. Klaman, O. Boss, O.D. Peroni, J.K. Kim, J.L. Martino, J.M. Zabolotny, N. Moghal, M. Lubkin, Y.B. Kim, A.H. Sharpe, A. Stricker-Krongrad, G.I. Shulman, B.G. Neel, B.B. Kahn, Increased energy expenditure, decreased adiposity, and tissue-specific insulin sensitivity in protein-tyrosine phosphatase 1B-deficient mice, Molecular and Cellular Biology 20 (2000) 5479–5489.
[17] S. Koren, I.G. Fantus, Inhibition of the protein tyrosine phosphatase PTP1B: potential therapy for obesity, insulin resistance and type-2 diabetes mellitus, Best Practice & Research Clinical Endocrinology & Metabolism 21 (2007) 621–640.
[18] A.B. Comeau, D.A. Critton, R. Page, C.T. Seto, A focused library of protein tyrosine phosphatase inhibitors, Journal of Medicinal Chemistry 53 (2010) 6768–6772.
[19] B.P. Kennedy, C. Ramachandran, Protein tyrosine phosphatase-1B in diabetes, Biochemical Pharmacology 60 (2000) 877–883.
[20] K. Shen, Y.F. Keng, L. Wu, X.L. Guo, D.S. Lawrence, Z.Y. Zhang, Acquisition of a specific and potent PTP1B inhibitor from a novel combinatorial library and screening procedure, Journal of Biological Chemistry 276 (2001) 47311–47319.
[21] C.O. Arregui, J. Balsamo, J. Lilien, Impaired integrin-mediated adhesion and signaling in fibroblasts expressing a dominant-negative mutant PTP1B, Journal of Cell Biology 143 (1998) 861–873.
[22] P. Pathre, C. Arregui, T. Wampler, I. Kue, T.C. Leung, J. Lilien, J. Balsamo, PTP1B regulates neurite extension mediated by cell–cell and cell–matrix adhesion molecules, Journal of Neuroscience Research 63 (2001) 143–150.
[23] G. Xu, C. Arregui, J. Lilien, J. Balsamo, PTP1B modulates the association of beta catenin with N-cadherin through binding to an adjacent and partially overlapping target site, Journal of Biological Chemistry 277 (2002) 49989–49997.
[24] J. Revuelta-Cervantes, R. Mayoral, S. Miranda, A. Gonzalez-Rodriguez, M. Fernandez, P. Martin-Sanz, A.M. Valverde, Protein tyrosine phosphatase 1B (PTP1B) deficiency accelerates hepatic regeneration in mice, American Journal of Pathology 178 (2011) 1591–1604.
[25] A.K. Greene, M. Puder, Partial hepatectomy in the mouse: technique and perioperative management, Journal of Investigative Surgery 16 (2003) 99–102.
[26] M. Piotto, K. Elbayed, J.-M. Wieruszeski, G. Lippens, Practical aspects of shimming a high resolution magic angle spinning probe, Journal of Magnetic Resonance 173 (2005) 84–89.
[27] M. Vinaixa, M.A. Rodriguez, A. Rull, R. Beltran, C. Blade, J. Brezmes, N. Canellas, J. Joven, X. Correig, Metabolomic assessment of the effect of dietary cholesterol in the progressive development of fatty liver disease, Journal of Proteome Research 9 (2010) 2527–2538.
[28] M.J. Albers, T.N. Butler, L. Rahwa, N. Bao, K.R. Keshari, M.G. Swanson, J. Kurhanewicz, Evaluation of the ERETIC method as an improved quantitative reference for H-1 HR-MAS spectroscopy of prostate tissue, Magnetic Resonance in Medicine 61 (2009) 525–532.
[29] N. Serkova, T.F. Fuller, J. Klawitter, C.E. Freise, C.U. Niemann, H-1-NMR-based metabolic signatures of mild and severe ischemia/reperfusion injury in rat kidney transplants, Kidney International 67 (2005) 1142–1151.
[30] T. Kishino, M. Tanno, H. Yamada, S. Saito, S. Matsumoto, Changes in liver fatty acid unsaturation after partial hepatectomy in the rat, Lipids 35 (2000) 445–452.
[31] H. Ishihara, K. Tamiyakoizumi, H. Kuriki, S. Yoshida, K. Kojima, Growth-associated changes in fatty-acid compositions of nuclear phospholipids of liver-cells, Biochimica Et Biophysica Acta 1084 (1991) 53–59.
[32] T. Shiina, T. Terano, J. Saito, Y. Tamura, S. Yoshida, Eicosapentaenoic acid and docosahexaenoic acid suppress proliferation of vascular smooth-muscle cells, Atherosclerosis 104 (1993) 95–103.
[33] Tj Delahunt, D. Rubinste, Accumulation and release of triglycerides by rat liver following partial hepatectomy, Journal of Lipid Research 11 (1970) 536.
[34] E.P. Newberry, S.M. Kennedy, Y. Xie, J. Luo, S.E. Stanley, C.F. Semenkovich, R.M. Crooke, M.J. Graham, N.O. Davidson, Altered hepatic triglyceride content after partial hepatectomy without impaired liver regeneration in multiple murine genetic models, Hepatology 48 (2008) 1097–1105.
[35] E. Shteyer, Y.J. Liao, L.J. Muglia, P.W. Hruz, D.A. Rudnick, Disruption of hepatic adipogenesis is associated with impaired liver regeneration in mice, Hepatology 40 (2004) 1322–1332.
[36] C.W. Sirey, M.S. Winters, M.M. Markiewski, J.D. Lambris, Partial hepatectomy induced liver proteome changes in mice, Proteomics 5 (2005) 318–325.
[37] Z.Z. Huang, Z. Mao, J. Cai, S.C. Lu, Changes in methionine adenosyltransferase during liver regeneration in the rat, American Journal of Physiology 275 (1998) G14–G21.
[38] J.M. Mato, S.C. Lu, Role of S-adenosyl-L-methionine in liver health and injury, Hepatology 45 (2007) 1306–1312.
[39] M.E. Bollard, N.R. Contel, T.M.D. Ebbels, L. Smith, O. Beckonert, G.H. Cantor, L. Lehman-McKeeman, E.C. Holmes, J.C. Lindon, J.K. Nicholson, H.C. Keun, NMR-based metabolic profiling identifies biomarkers of liver regeneration following partial hepatectomy in the rat, Journal of Proteome Research 9 (2010) 59–69.
[40] L.X. Chen, Y. Zeng, H.P. Yang, T.D. Lee, S.W. French, F.J. Corrales, E.R. Garcia-Trevijano, M.A. Avila, J.M. Mato, S.C. Lu, Impaired liver regeneration in mice lacking methionine adenosyltransferase 1A, FASEB Journal 18 (2004) 914.
[41] P.J. O'Brien, M.R. Slaughter, S.R. Polley, K. Kramer, Advantages of glutamate dehydrogenase as a blood biomarker of acute hepatic injury in rats, Laboratory Animals 36 (2002) 313–321.

OPEN ACCESS Freely available online

PLoS one

Metabolomics Approach for Analyzing the Effects of Exercise in Subjects with Type 1 Diabetes Mellitus

Laura Brugnara[1,5], Maria Vinaixa[2,5], Serafín Murillo[1,5], Sara Samino[2,3], Miguel Angel Rodriguez[2,5], Antoni Beltran[2,3], Carles Lerin[1,5], Gareth Davison[4], Xavier Correig[2,3,5], Anna Novials[1,5]*

1 Department of Endocrinology, Institut d'Investigacions Biomèdiques August Pi i Sunyer (IDIBAPS), Hospital Clínic de Barcelona, Barcelona, Spain, 2 Metabolomics Platform, Universitat Rovira i Virgili, Tarragona, Spain, 3 Institut d'Investigació Sanitària Pere Virgili (IISPV), Reus, Spain, 4 Sport and Exercise Sciences Research Institute, University of Ulster, Newtownabbey, Northern Ireland, United Kingdom, 5 Spanish Biomedical Research Centre in Diabetes and Associated Metabolic Disorders (CIBERDEM), Barcelona, Spain

Abstract

The beneficial effects of exercise in patients with type 1 diabetes (T1D) are not fully proven, given that it may occasionally induce acute metabolic disturbances. Indeed, the metabolic disturbances associated with sustained exercise may lead to worsening control unless great care is taken to adjust carbohydrate intake and insulin dosage. In this work, pre- and post-exercise metabolites were analyzed using a ^{1}H-NMR and GC-MS untargeted metabolomics approach assayed in serum. We studied ten men with T1D and eleven controls matched for age, body mass index, body fat composition, and cardiorespiratory capacity, participated in the study. The participants performed 30 minutes of exercise on a cycle-ergometer at 80% VO_2max. In response to exercise, both groups had increased concentrations of gluconeogenic precursors (alanine and lactate) and tricarboxylic acid cycle intermediates (citrate, malate, fumarate and succinate). The T1D group, however, showed attenuation in the response of these metabolites to exercise. Conversely to T1D, the control group also presented increases in α-ketoglutarate, alpha-ketoisocaproic acid, and lipolysis products (glycerol and oleic and linoleic acids), as well as a reduction in branched chain amino acids (valine and leucine) determinations. The T1D patients presented a blunted metabolic response to acute exercise as compared to controls. This attenuated response may interfere in the healthy performance or fitness of T1D patients, something that further studies should elucidate.

Citation: Brugnara L, Vinaixa M, Murillo S, Samino S, Rodriguez MA, et al. (2012) Metabolomics Approach for Analyzing the Effects of Exercise in Subjects with Type 1 Diabetes Mellitus. PLoS ONE 7(7): e40600. doi:10.1371/journal.pone.0040600

Editor: Rocio I. Pereira, University of Colorado Denver, United States of America

Received January 25, 2012; **Accepted** June 11, 2012; **Published** July 11, 2012

Funding: This work was supported by CIBERDEM – Spanish Biomedical Research Centre in Diabetes and Associated Metabolic Disorders (ISCIII, Ministerio de Ciencia e Innovación); METADIAB (Metabolism and Diabetes Project) 12-12-2009. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: anovials@clinic.ub.es

Introduction

Type 1 diabetes mellitus (T1D) is a lifelong metabolic disorder of usual acute onset in children, adolescents and young adult people. Over time, micro and macro vascular co-morbidities develop in patients with T1D which are closely related to metabolic control [1]. In addition to these complications, the management of T1D is particularly complex and challenging. It is well known that, in order to maintain optimal glycemic control, T1D patients need accurate administration of insulin coordinated with a balanced diet and an adequate level of physical activity.

Exercise plays a crucial role in the prevention and treatment of several chronic diseases, including glucose intolerance states, type 2 diabetes [2–4] and diseases of the cardiovascular system [5,6]. Moreover, it has been demonstrated that exercise improves the quality of life in the general population [7]. However, the beneficial effects of exercise in patients with T1D are not fully proven, given that exercise may occasionally induce acute metabolic disturbances, mainly related to insulin treatment. Nevertheless, children and adolescents with T1D are encouraged to exercise regularly as a means of improving social integration and cardiovascular health [8]. Thus, a better understanding of the effects of exercise on the metabolic response in T1D patients will allow clinicians to prescribe exercise to their patients with greater clarity.

Metabolomics enables the systematic assessment of the abundant changes of low molecular weight compounds present in biological samples, using high-throughput sample analysis techniques (GC-MS, NMR or HPLC-MS) and computer-assisted multivariate pattern-recognition techniques [9]. Metabolomics is enriching our current understanding of both the physiologic and physiopathologic processes underlying diabetes mellitus [10–12]. Moreover, recent metabolomic-based studies have described the first metabolic signatures of exercise in human plasma [13–15]. For example, Lewis et al [13] described the metabolic changes in tricarboxylic acid cycle, fatty acid oxidation and lypolisis in the plasma of healthy subjects exposed to different intensities and durations of exercise.

Most of the *in vivo* studies investigating the metabolic pathways of T1D have been performed under strictly controlled conditions using hyperinsulinemic euglycemic clamp techniques [16,17] or in situations of insulin withdrawal [12]. Although these studies have provided invaluable new insights into the metabolic disturbances present in T1D, the experimental conditions used are dissimilar to everyday life. To the best of our knowledge, no study to date has

Annex

applied a metabolomics approach prior to and following exercise in subjects with T1D. We hypothesize that (a) an acute bout of exercise will result in changes in the systemic metabolic profile and that (b) these parameters will be different in patients with T1D in comparison to healthy controls. The aim of this study is to analyze the metabolic changes induced by a short-term session of acute exercise performed by T1D patients and their corresponding non-diabetic counterparts. A comprehensive ^{1}H-NMR and GC-MS untargeted metabolomics approach was applied to serum samples taken from all participants.

Methods

Participants

Ten recreationally active male patients with T1D, recruited by the Department of Endocrinology (Hospital Clinic, Barcelona), and eleven non-diabetic controls matched for sex, age, body mass index (BMI) and similar physical activity, recruited from a research institute (IDIBAPS, Barcelona), were enrolled in the study.

The T1D patients participating in the study had diabetes for a total of 14±8.4 years, undetectable C-peptide levels and good glycemic control, as determined by glycated hemoglobin A1c. Total body composition was measured by densitometry using DXA (Lunar iDXA body composition, GE Healthcare). Patients with chronic complications related to diabetes were excluded. All patients presented microalbuminuria values below 30 mg/L, normal retinal exam by direct and indirect retinoscopy, normal peripheral neurologic evaluation by clinical exploration and biothesiometry (Bio-thesiometer, Bio-Medical Instrument Company, Newbury, OH, U.S.), and normal, resting 12-lead electrocardiogram (ECG) and normal exercise testing by upright cycle-ergometer (25 W/3 min) [18]. At the time of testing, none of the T1D participants were taking any form of prescription medication, except for long-acting basal insulin analogue glargine (Sanofi-Aventis, U.S.) and fast-acting insulin analogue aspart (Novo Nordisk, Denmark).

Experimental Procedures

All subjects were required to visit the Diabetes and Exercise Research Unit of the Hospital Clinic on two separate occasions. On the first visit, all subjects were fully briefed and familiarized with the experimental procedures. Baseline clinical characteristics such as height, weight, BMI and total and fat body composition where also obtained, and each subject was required to complete an evaluation of current physical activity using the International Physical Activity Questionnaire [19]. Maximal oxygen uptake (VO_2max) was determined by using a maximal progressive incremental exercise test on a friction-braked cycle-ergometer (Monark 828E, Monark Sweden). After a 3 min warm-up period at a power output of 25-W, workload was increased by 25-W every minute until exhaustion. Oxygen uptake was monitored during exercise using a computerized, open circuit gas-collection system (Vmax Spectra, version v12.0, Sensor Medics Corp, VIASYS Healthcare Inc, Yorba Linda, CA, U.S.), and VO_2max was determined at the point of highest oxygen consumption over a 15-s period. VO_2max was confirmed using established physiological criteria, including a respiratory exchange ratio above 1.15, oxygen uptake reaching a plateau despite an increased work rate, and a heart rate near 95% of the age-predicted maximum value.

On the second visit, conducted at the same time (8:30am) a week later, subjects performed an acute bout of 30 minutes of intense exercise at 80% VO_2max (individually calculated during the preliminary session) on a cycle-ergometer. Subjects performed 3 to 6 minutes of warm-up until achieving a fixed cardiac frequency. Prior to exercise, all participants fasted overnight for a 12 hour period, following a balanced meal consisting of approximately 55% carbohydrates, 30% fat, and 15% proteins. Furthermore, T1D patients had received their last short-acting insulin injection before dinner at 8 p.m. and their long-acting insulin injection at 10 p.m. All subjects were asked to avoid strenuous exercise and alcohol consumption the day prior to the acute exercise protocol.

Blood determinations. Fasting blood samples were obtained before and after the short-term intensive exercise intervention. Glycemia (glucose-oxidase method, Advia 2400 Siemens Diagnostics, Deerfield, IL, U.S.) and insulinemia (quimioluminiscent method, Siemens Healthcare Diagnostics, Tarrytown, NY, U.S.) were determined in serum samples. For the metabolomic measurements, serum was obtained once blood had been allowed to clot at room temperature for 30 min and after centrifugation at 4°C at 5000 rpm for 10 min. Samples were kept at - 80°C until further metabolomic analysis.

Serum ^{1}H-NMR metabolomics. Serum samples were thawed, vortexed and allowed to stand for 10 min prior to NMR analysis. For NMR measurements 430 µl of serum were transferred into 5 mm NMR tubes. A double tube system was used: an internal tube (o.d. 2 mm, supported by a Teflon adapter) containing the reference substance (sodium 3-trimethylsilyl [2, 2, 3, 3-d4] propionate (TSP) 9.9 mmol/l, $MnSO_4$ 0.47 mmol/l in 99.9% D_2O) was placed coaxially into the NMR sample tube (o.d. 5 mm). This double tube system was kept at 4°C in the sample changer until analysis was performed. Spectra were acquired at a ^{1}H observation frequency of 600.20 MHz at a temperature of 300 K using an Avance III-600 Bruker spectrometer equipped with an inverse TCI 5 mm cryoprobe®. The Carr-Purcell-Meiboom-Gill (cpmg, spin-spin T_2 relaxation filter) pulse sequence with a fixed spin-spin relaxation delay of 200 ms was applied to acquire ^{1}H-NMR spectra for all serum samples, in order to minimize the broad signals arising from lipoprotein and albumin in the NMR spectra. For each sample, 128 transients were collected into 32 K data points using a spectral width of 12 kHz with a relaxation delay of 2 s and an acquisition time of 1.36 s. A line-broadening function of 0.3 Hz was applied to all spectra prior to Fourier transform.

Serum GC-MS metabolomics. A second aliquot of serum (100µL) was used for GC-MS analysis according to Agilent's specifications [20]. Each aliquot was spiked with 20 µl internal standard solution (1 µg.µL-1 succinic-d4 acid; Sigma-Aldrich). After protein precipitation using 900 µl of cold methanol/water (8:1 v/v), samples were centrifuged for 10 minutes at 4°C. 200 µL of the supernatant were transferred to a GC autosampler vial and spiked with 20 µl of myristic acid-d27 (Sigma Aldrich), used as the internal standard for retention time lock (RTL system provided in Agilent's ChemStation Software), and lyophilized overnight (Lyo-trap freeze dryer). Samples were methoximated by incubating lyophilized serum residues in 50 µl of methoxyamine in pyridine (0.3 µg/µL) for 16 hours at room temperature. Silylation was subsequently done using 30 µL of N-methyl-N-trimethylsilyltri-fluoroacetamide with 1%trimethylchlorosilane (MSTFA +1% TMCS, Sigma) for 1 hour at room temperature. Samples were automatically injected into a GC–MS system (HP 6890 Series gas chromatograph coupled to a mass selective detector model 5973) equipped with a J&W Scientific DB 5-MS+DG stationary phase column (30 m × 0.25 mm i.d., 0.1 µm film) (Agilent Technologies). The injector temperature was set at 250°C, and the helium carrier flow rate was kept constant at 1.1 mL/min. The column temperature was held at 60°C for 1 min, then increased to 325°C

at a rate of 10°C/min and held at 325°C for 10 min. The detector operated in the electron impact ionization mode (70 eV) and mass spectra were recorded after a solvent delay of 4 min with 2.46 scans per second (mass scanning range of m/z 50–600; threshold abundance value of 50 counts). The source temperature and quadrupole temperature were 230 and 150°C, respectively.

Ethics

Written informed consent was obtained from all subjects prior to participation. The experimental protocol was approved by the Research and Ethics committees of the Hospital Clinic de Barcelona, in accordance with the Declaration of Helsinki.

Data Analysis and Statistical Methods

The acquired CPMG ^{1}H-NMR spectra were phased, baseline-corrected and referenced to the chemical shift of the α-glucose anomeric proton doublet at 5.23 ppm. Pure standards compound reference in BBioref AMIX (Bruker) was used; HMDB and Chenomx databases were used for metabolite identification. After baseline correction, intensities of each ^{1}H-NMR region identified in the CPMG 1D-NMR spectra were integrated using the AMIX 3.8 software package (Bruker, GmBH). Each region was normalized to the ERETIC (Electronic REference To access *In vivo* Concentrations) signal [21].

Raw GC/MS files were exported into the platform-independent netCDF (*.cdf) and loaded into XCMS software (version 1.6.1) based on R-program version 2.4.0 (R-Foundation for statistical computing, www.Rproject.org), where peak peaking, integration and alignment in the time domain were performed. Integrated intensities of each m/z-retention time pair (MZRT) were obtained for each one of the samples used in the study. These intensities were normalized to internal standard succinic acid-d4. The AMDIS program (Automated Mass Spectral Deconvolution and Identification System, National Institute of Standards and Technology, Gathersburg, MD, U.S.A.) was run for peak annotation, and both the Fiehn GC/MS Metabolomics RTL Library and NIST mass spectral databases were used for identification. Table S1 shows detailed identification parameters for both GC-MS and ^{1}H-NMR determined metabolites.

Baseline metabolic differences between T1D and control groups were evaluated using the Mann-Witney U run test. Effects of exercise intervention on the independent control and T1D groups were assessed using the Wilcoxon exact rank sum tests. Repeated measures ANOVA were used to determine diabetes×exercise interactions. A statistically significant interaction indicates that control and T1D responded differently to the acute exercise protocol for a given metabolite. To account for multiple testing, q-values were computed for all systematic univariate tests outlined above by applying the FDR (False Discovery Rate) procedure described by Storey et al [22]. In all cases statistical significance was set at $q \leq 0.1$.Data (pre-) processing, data analysis, and statistical calculations were performed with Matlab (Matlab version 6.5.1, Release 13, The Mathworks, 2003).

Results

Clinical and Biochemical Baseline Characteristics

Anthropometric and fitness data are summarized in Table 1. In ten patients with T1D and eleven controls matched for age, height, weight, and body mass index (BMI), no differences were found in relation to the percentage of body fat composition and cardio respiratory capacity evaluated by VO_2max.

At baseline (Table 2), subjects with T1D presented higher glucose and insulin levels than controls. Moreover, untargeted ^{1}H-NMR and GC-MS metabolomic profiling results showed elevated levels of the tricarboxylic acid cycle intermediates (TCAIs) malate and citrate in T1D. Glycerol was also increased in the T1D group, whereas lysine levels were significantly lower as compared to the control group.

Table 1. Clinical characteristics of T1D patients and control population.

	Control	T1D	p-values
Subjects	11	10	ns
Age (years)	32.5±8.8	35.1±8.4	ns
Evolution of diabetes (years)	–	14±8.4	–
Height (m)	1.76±0.06	1.76±0.05	ns
Weight (kg)	75.9±8.6	75.6±5.8	ns
BMI (kg/m^2)	24.7±2.6	24.3±1.7	ns
Fat percentage (%, by DXA)	23.9±5.7	21.7±6.5	ns
IPAQ (METs min/week)	2550±995	2630± 241	ns
VO_2max (mL/kg/min)	34±9.1	35±6.5	ns
Units of insulin glargine	–	31±7.9	–

Values are reported as mean values ± SD.
ns: not significant.
doi:10.1371/journal.pone.0040600.t001

Metabolic Changes Induced by Short-term Intensive Exercise

Figure 1A shows a comparison of the mean net percent variation following exercise in the individual T1D and control groups for glucose and insulin. After 30 minutes of intensive exercise, glucose levels varied differently according to diabetic condition. The T1D group, hyperglycaemic at baseline, showed significant glucose depletion with acute exercise unlike their control counterparts. Insulin levels also varied differently with the exercise according to diabetic condition. While insulin levels rose

Table 2. Baseline analytical differences of T1D patients and control population.

	Control (n = 11)	T1D (n = 10)	q-values
Biochemical determinations			
Glucose (mg/dl)	90.27±2.25	202.7±24.36	5.52×10^{-4}
Insulin (UI/L)	6.96±1.31	18.62±4.64	0.022
C-peptide (ng/ml)	–	undetectable	–
Glycated hemoglobin (%)	–	6.9±1	–
Metabolomics analysis (arbitrary units)			
Lysine	0.039±0.0036	0.025±0.0029	0.023
Glycerol	0.039±0.0022	0.045±0.0039	0.064
Citrate	0.006±0.0004	0.008±0.0005	0.052
Malate	0.001±0.0001	0.002±0.0002	0.083

Values are reported as mean values ±SEM. Selected quantitative ions relative to internal standard areas are used in the case of GC-MS measurements. Selective ^{1}H-NMR regions relative to ERETIC digital signals are used in the case of NMR measurements. Two-sided p-values are calculated using Mann-Witney test. Statistical significance was set as $q<0.1$.
doi:10.1371/journal.pone.0040600.t002

significantly with exercise in the T1D group, control exercisers showed a net decrease of insulin, which did not account for statistical significance. Although after short-term acute exercise both T1D and control groups showed a significant net increase in circulating levels of gluconeogenic precursors (alanine and lactate), this increase was less pronounced in the T1D group (Figure 1B). In addition to alanine and lactate, pyruvate increased significantly with exercise in the control group but not in the T1D group.

A significant enrichment in TCA cycle intermediates citrate, malate, fumarate, and succinate was observed after acute exercise in the peripheral blood of both control and T1D groups (Figure 2). However, α-ketoglutarate levels significantly increased only in the case of control exercisers. In general terms, there was less accumulation of TCA cycle intermediates in serum in the T1D group. No changes in other TCAIs were identified (isocitrate, succinil-coA and oxalacetate), nor was any interaction observed between the groups of T1D and controls in the metabolites that increased following exercise.

Exercise significantly increased glycerol and oleic and linoleic acid concentrations in the control group exclusively. This effect was attenuated in the T1D group (Figure 3A). There was no interactive effect observed between the T1D and control groups in these products of lipolysis. Branched chain amino acids (valine and leucine) were lower following short-term acute exercise in the control group (Figure 3B). These changes were paralleled by significantly increased levels of alpha-ketoisocaproic acid (2-KIC) in the same group. This increase was shown to be a diabetes-dependent trend (p-interaction diabetes×exercise = 0.05) but not statistically significant after FDR correction at the established conventional significance level ($q \leq 0.1$). Worth mentioning is lysine, which also showed a diabetes-dependent effect with the exercise.

Metabolite identification parameters are presented in supporting table (Table S1).

Discussion

The principal aim of this study was to analyze the metabolomic profile at rest and after a short period of intense exercise in patients affected by T1D to provide a comprehensive insight into the physiological effects of exercise on this particular population. Based on serum sample analysis, we have compared the metabolic response to short term acute exercise in T1D and healthy subjects using an untargeted metabolomics approach (GC-MS and ^{1}H-NMR). Our findings revealed similar metabolic events in T1D patients and their control matched exercisers, although the T1D patients showed an attenuation of overall metabolic response after intense short-term exercise.

As it is commonly known, in order to increase energy supply during intense short-term exercise, glycogen breakdown is induced

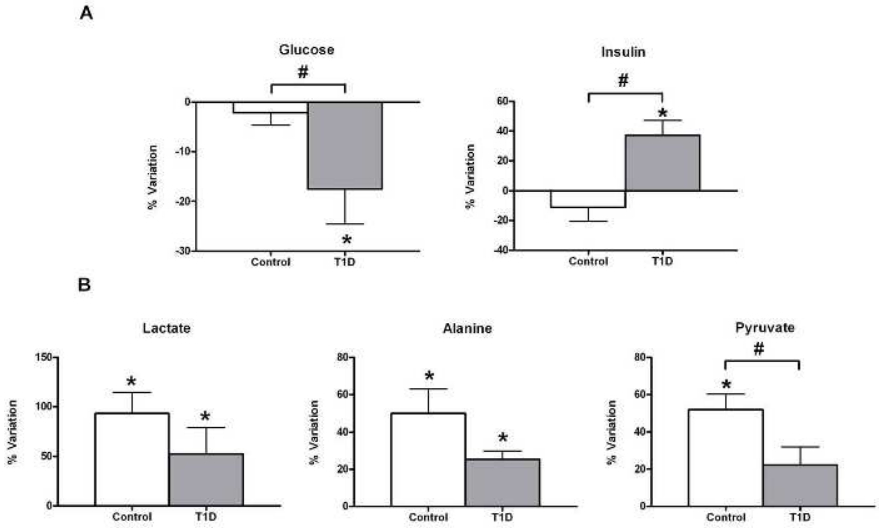


Figure 1. Relative changes in insulin and glucose and in gluconeogenic precursors in response to acute exercise. Relative changes in insulin and glucose (A) and in gluconeogenic precursors (B) in response to 30 minutes of acute exercise (80% VO_2max). Percentage of variation was calculated for each individual as the levels of a certain metabolite after exercise minus the levels of the same metabolite prior to exercise relative to the former. Data are shown as mean±sem of net percent variation for T1D and control groups. A positive value of percentage of variation indicates that metabolic levels have increased in mean with exercise, whereas a negative mean denotes the opposite. *Indicates a significant variation in metabolic levels with exercise (Wilcoxon rank-summed test for the comparison of a particular metabolite level prior to and after exercise in the independent T1D and control exercisers, $q<0.1$). #Indicates a significant diabetes×exercise interaction for a particular metabolite (Repeated-measures ANOVA, $q<0.1$). Insulin and glucose data correspond to biochemical measurements, whereas lactate, alanine, and pyruvate were evaluated in ^{1}H-NMR spectra according to Table S1.
doi:10.1371/journal.pone.0040600.g001

Annex

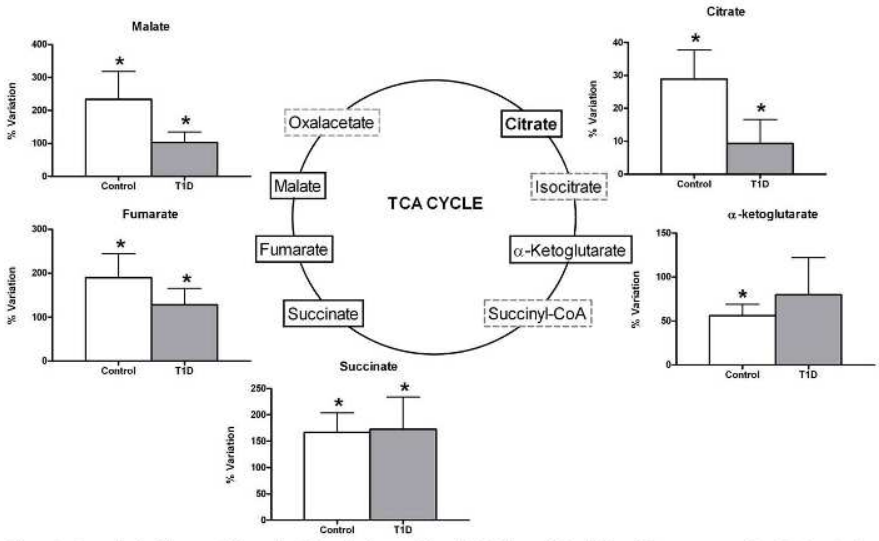


Figure 2. General significant enrichment in TCA cycle intermediates (TCAIs) in peripheral blood after acute exercise. Monitored using GC-MS (malate, fumarate, α-ketoglutarate) and NMR (citrate and succinate). Data are mean±sem of net percent variation with exercise. *Indicates a significant variation in metabolic levels with exercise (Wilcoxon rank-summed test for the comparison of a particular metabolite level prior to and after exercise in the independent T1D and control exercisers, $q<0.1$).
doi:10.1371/journal.pone.0040600.g002

to provide the substrate for activating anaerobic glycolysis, resulting in the accumulation of pyruvate and lactate in plasma. Recently, Lewis et al [13] confirmed these results by using LC-MS-based metabolomics in a non-diabetic population. Our data, using a different metabolomic approach based on ^{1}H-NMR and GC-MS techniques, showed that the non-diabetic group presented an increase in serum lactate and pyruvate concentration after exercise, indicating that glycolysis was activated, as previously reported by other groups [13,23]. However, the T1D patients elicited only an increase in lactate levels having blunted the pyruvate response, suggesting that exercise induces a reduced activation of glycogenolysis and glycolysis in the T1D group as compared to the control group. Given the inhibitory effect of insulin on glycogen breakdown, the higher insulin levels observed in the T1D group could explain in part the attenuated glycogenolytic response.

Our analysis demonstrated a significant increase in TCAIs (malate, citrate, succinate, fumarate and α-ketoglutarate) in the control subjects in response to exercise. These results are in accordance with a previous metabolomics-based study investigating the plasma signature of exercise [13], in which accumulated levels of malate, fumarate and succinate were reported following 60 minutes of exercise in a healthy cohort. Other studies have demonstrated that an increase in the total concentration of TCAIs is necessary to enhance and maintain TCA cycle flux during strenuous exercise [24]. In addition, a positive correlation has been demonstrated between the total concentration of TCAIs and the estimated TCA cycle flux in human skeletal muscle taken from muscle biopsies during exercise [25]. In the case of patients with T1D, our results also pointed to a less significant accumulation of TCAIs after exercise, which might also compromise the TCA flux rates. Given that many of the reactions that lead to a net influx of TCAIs are directly or indirectly dependent on the level of pyruvate and that elevated concentrations of pyruvate appear to be necessary for the anaplerosis pathway [25], we suggest that the insufficient increase in pyruvate concentrations in T1D subjects might have an impact on TCA cycle replenishment. In this sense, our findings of an attenuated enrichment of TCAIs in the serum of T1D patients in response to a short period of intense exercise indicate that the activation of the TCA cycle flux rates might be affected. This suggests that the T1D group has a compromised oxidative aerobic system as compared to the control group, as the T1D group did not show a significant turnover of TCA metabolites.

Concerning lipolysis, our data demonstrated that the healthy controls showed an increase in free fatty acids and glycerol in response to exercise, however, this response was attenuated in the T1D group. The increase in lipolysis after acute exercise had been previously reported in healthy individuals [13,26]. Under conditions of hyperinsulinemia induced by clamp techniques, a suppression of the intramuscular and subcutaneous adipose tissue lipolysis was demonstrated in healthy volunteers [27,28]. In another study performed in obese men, a delay in the lipolytic activation was observed together with an increase in plasma insulin levels following 30 minutes of acute resistance exercise, a finding that the authors attributed to an increase in insulin levels

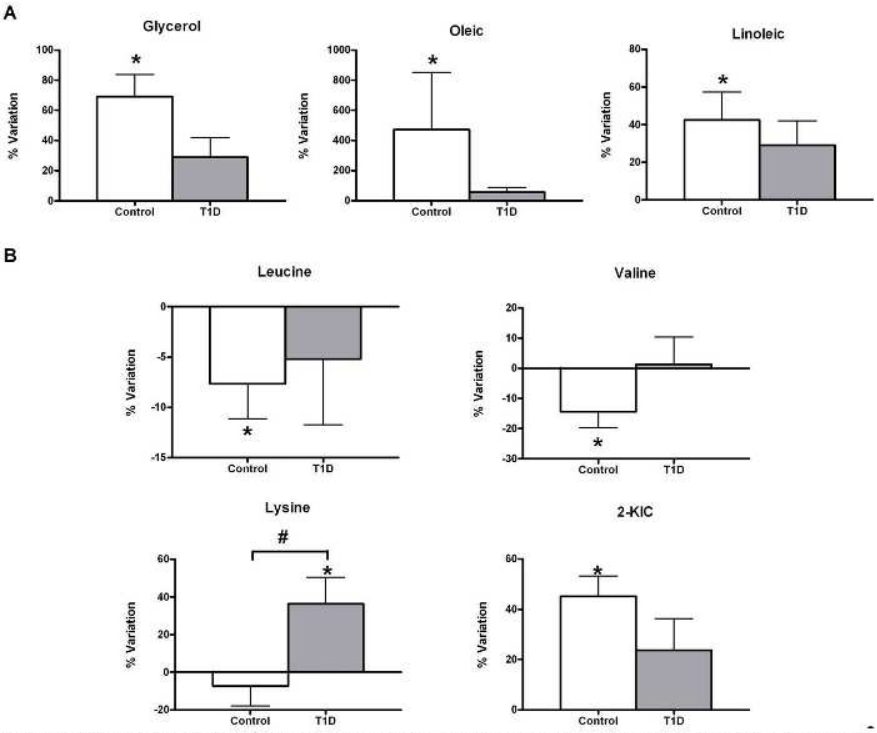


Figure 3. Relative changes in lypolysis (A) and BCAA metabolism (B) with acute exercise. Data are mean±sem of net percent variation. *Indicates a significant variation in metabolic levels with exercise (Wilcoxon rank-summed test for the comparison of a particular metabolite level prior to and after exercise in the independent T1D and control exercisers, q<0.1). #Indicates a significant diabetes×exercise interaction for a particular metabolite (Repeated-measures ANOVA, q<0.1).
doi:10.1371/journal.pone.0040600.g003

[26]. The increased insulin levels found in the T1D group in our study could have a role in the attenuated lipolytic action observed following exercise.

Our results demonstrated that control exercisers presented a significant reduction in leucine levels. It is well established that exercise increases energy expenditure, resulting in the promotion of amino acid catabolism in general and, in particular, in the oxidation of branched chain amino acids, mainly leucine [13,29]. In parallel to the reduction of leucine, we observed an increase in circulating levels of 2-KIC, which is the first step in leucine degradation. This increase in 2-KIC is in concordance with a previous reported study on the metabolomic profiling performed on urine samples of men following acute exercise [15]. Our results suggest that the effect induced by exercise on protein catabolism was attenuated in the diabetic patients, in agreement with other authors who have previously demonstrated a decrease in protein catabolism in T1D patients [17]. Moreover, there are evidences confirming that elevated insulin levels induced by infusion in healthy men promote muscle protein anabolism by inhibiting protein breakdown [30] and elicit the ability to stimulate glucose uptake and alanine transport, thus suppressing protein degradation in skeletal muscle [31]. Taking all these data into consideration, we propose that the high insulinemia levels induced by exogenous insulin administration in our T1D group could be the cause of reduced protein breakdown, as demonstrated by the minor alterations observed in the levels of leucine and 2-KIC following exercise.

Of special note, we detected an increase in insulin serum levels in all the T1D patients after 30-minute of acute exercise. This increase may explain in part the attenuation in the metabolomic response of all energetic substrates. Pharmacokinetic studies have shown that the peak action of insulin glargine is usually within the first 3 or 4 hours after injection [32], and in our protocol, samples were taken 10 hours later. One possible explanation for the increased insulin concentration detected in the serum of our T1D group could be that intense exercise induces the lipolysis of

subcutaneous adipose tissue where insulin is stored and rapidly released into the circulation. In line with this hypothesis, Davison et al [33] previously reported an exercise-induced lipolysis effect by monitoring the release of liposoluble vitamins from the subcutaneous tissue into the bloodstream, an effect which may be responsible for the increased insulin levels observed in our T1D patients.

Although we speculate that high insulin levels induced by exogenous treatment may be responsible for the attenuated response of metabolites to acute exercise, alternative explanations must also be considered. For example, the possible presence of insulin resistance (IR), an important condition described in T1D patients, cannot be ignored. Some authors have considered that supra-physiologic levels of exogenous insulin [34] and hyperglycemia per se [35] could be responsible for IR in T1D patients. Insulin, not the rate of glucose disposal per se, regulates glycogen synthesis, meaning that the low level of glycogen synthase activity found in insulin-resistant states is a consequence of impaired insulin action, rather than reduced glucose disposal. Nevertheless, this assumption has been challenged in a more recent study in which adult patients with T1D exhibited both impaired glucose utilization and impaired insulin-induced non-esterified fatty acid suppression [36]. In addition, these patients showed IR in hepatic and skeletal muscle tissue, despite good glycemic control [37]. Another report [38] demonstrated that T1D adolescents had significantly impaired functional exercise capacity and decreased insulin sensitivity as compared to non-diabetic adolescents. In spite of their IR and reduced cardiovascular fitness, T1D youth showed paradoxically normal intramyocellular lipid content (IMCL). This finding contradicts the previously well- established theory that IMCL accumulation is a marker for insulin resistance in both T1D and T2D [39,40].

Metabolic flexibility defined as the ability to switch from fat to carbohydrate oxidation is usually impaired during a hyperinsulinemic clamp in insulin-resistant subjects. Thus, the phenomena of metabolic inflexibility mainly described in T2D and other insulin-resistant states could explain some of the alterations occurring in the machinery of lipid and glucose metabolism [41,42]. In addition, the inability to modify fuel oxidation in response to changes in nutrient availability has been implicated in the accumulation of intramyocellular lipids as well as in insulin resistance. Exercise represents a paradigm which requires a highly regulated coordination between fuel supply and oxidative machinery. The assumption that metabolic inflexibility may affect energy metabolism under exercise conditions in T1D patients remains to be elucidated.

Several studies involving T2D patients and their offspring have demonstrated the presence of mitochondrial dysfunction [43,44]. In T1D patients, mitochondrial dysfunction has also been described, contributing to abnormalities in the TCA cycle and fatty acid metabolism. A recent report [45] verified that the mitochondrial capacity of untrained women with T1D correlates positively with glycemic control. The hypothesis that mitochondrial abnormalities may be a primary cause of metabolic inflexibility and insulin resistance has been offered. Significant differences in mitochondrial number, structure and function have been described between insulin-resistant and insulin-sensitive subjects, but the causal link still remains unknown [46].

Although our study was not designed to analyze IR or metabolic flexibility, these factors might have influenced the metabolomic spectrum described in our T1D patients. In addition, the presence of hyperglucagonemia and impaired glucagon counterregulation, as reported in several studies [47], could affect lipolysis, gluconeogenesis and protein metabolism. Finally, other situations not present in our patients should be taken into account, such as autonomic dysfunction, which influences hypolitic responsiveness [48], cardiac dysfunction, which has been documented in T1D patients as affecting exercise responses [49,50], and abnormal blood flow during exercise, which could affect muscular vascular function, contributing toward metabolic disturbances [51].

In summary, we report that T1D patients have an attenuated metabolic response as compared to their healthy control counterparts after a short period of acute, intense exercise. We speculate that exercise could mobilize the subcutaneous exogenous insulin depot in adipose tissue. Furthermore, our data suggest that high insulinemia levels might play a role in the attenuated response in lipolysis, proteolysis, glycogenolysis, and oxidative metabolism observed in T1D patients following exercise. Whether the attenuation of metabolic response to acute and intense exercise might interfere in the training performance of T1D patients remains to be elucidated, and additional studies are required.

Supporting Information

Table S1 Metabolite identification parameters. ^{1}H-NMR: proton nuclear magnetic resonance spectroscopy; **δ**: chemical shift; **GC-MS**: gas chromatography-mass spectrometry; **RT:** retention time
(DOC)

Acknowledgments

We are grateful to Kimberly Katte for the editorial assistance.

Author Contributions

Conceived and designed the experiments: LB MV SM XC AN. Performed the experiments: LB SM MV SS MAR AB. Analyzed the data: MV LB SS MAR GD CL. Contributed reagents/materials/analysis tools: LB SM MV SS MAR AB. Wrote the paper: LB MV AN.

References

1. DCCT/EDIC - Diabetes Control and Complications Trial/Epidemiology of Diabetes Interventions and Complications (DCCT/EDIC) Study Research Group (2005) Intensive diabetes treatment and cardiovascular disease in patients with type 1 diabetes. N Engl J Med 353: 2643–2653.
2. DPP - Diabetes Prevention Program Research Group (2002) Reduction in the incidence of type 2 diabetes with lifestyle intervention or metformin. N Engl J Med 346: 393–403.
3. Pan XR, Li GW, Hu YH, Wang JX, Yang WY, et al. (1997) Effects of diet and exercise in preventing NIDDM in people with impaired glucose tolerance. The Da Qing IGT and Diabetes Study. Diabetes Care 20: 537–544.
4. Tuomilehto J, Lindström J, Eriksson JG, Valle TT, Hämäläinen H, et al. (2001) Prevention of type 2 diabetes by changes in lifestyle among subjects with impaired glucose tolerance. N Engl J Med 344: 1343–1350.
5. Thompson PD, Buchner D, Piña IL, Balady GJ, Williams MA, et al. (2003) Exercise and physical activity in the prevention and treatment of atherosclerotic cardiovascular disease - a statement from the Council on Clinical Cardiology (Subcommittee on Exercise, Rehabilitation, and Prevention) and the Council on Nutrition, Physical Activity, and Metabolism (Subcommittee on Physical Activity). Circulation 107: 3109–3116.
6. Leon AS, Rice T, Mandel S, Després JP, Bergeron J, et al. (2000) Blood lipid response to 20 weeks of supervised exercise in a large biracial population: the HERITAGE Family Study. Metabolism 49: 513–520.
7. Haskell WL, Lee IM, Pate RR, Powell KE, Blair SN, et al. (2007) Physical activity and public health: updated recommendation for adults. From the American College of Sports Medicine and the American Heart Association. Circulation 116: 1081–1093.
8. Valerio G, Spagnuolo ML, Lombardi F, Spadaro R, Sinno M, et al. (2007) Physical activity and sports participation in children and adolescents with type 1 diabetes mellitus. Nutr Metab Cardiovasc Dis 17: 376–382.

Annex

9. Lindon JNJ, Holmes E (2007) The handbook of metabonomics and metabolomics. 1 ed. UK: Elsevier.
10. Makinen VP, Soininen P, Forsblom C, Parkkonen M, Ingman P, et al. (2008) Diagnosing diabetic nephropathy by 1H-NMR metabonomics of serum. Magn Reson Mater Phy 19: 281–296.
11. Bain JR, Stevens RD, Wenner BR, Ilkayeva O, Muoio DM, et al. (2009) Metabolomics applies to diabetes research. Moving from information to knowledge. Diabetes, 58: 2429–2443.
12. Lanza IR, Zhang S, Ward LE, Karakelides H, Raftery D, et al. (2010) Quantitative metabolomics by 1H-NMR and LC-MS/MS confirms altered metabolic pathways in diabetes. PLoS ONE 5: e10538, 1–10.
13. Farrell L, Wood MJ, Martinovic M, Arany Z, et al. (2010) Metabolic signatures of exercise in human plasma. Sci Transl Med 2: 33ra37. doi:10.1126/scitranslmed.3001006.
14. Enea C, Seguin F, Petitpas-Mulliez J, Boildieu N, Boisseau N, et al. (2010) ^{1}H NMR-based metabolomics approach for exploring urinary metabolome modifications after acute and chronic physical exercise. Anal Bioanal Chem 396: 1167–1176.
15. Peelvlianis A, Kostidis S, Saraslanidis P, Petridou A, Tsalis G, et al. (2010) 1H NMR-Based metabonomic investigation of the effect of two different exercise sessions on the metabolic fingerprint of human urine. J Proteome Res 9: 6405–6516.
16. Zhang S, Zheng C, Lanza IR, Nair S, Raftery D, et al. (2009) Interdependence of signal processing and analysis of urine 1H NMR spectra for metabolomic profiling. Anal Chem. 81: 6080–6088.
17. Devlin J, Scrimgeour A, Brodsky I, Fuller S (1994) Decreased protein catabolism after exercise in subjects with IDDM. Diabetologia 37: 358–364.
18. Palmieri V, Pezzullo S, Arezzi E, D'Andrea C, Cassese S, et al. (2008) Cycle-ergometry stress testing and use of chronotropic reserve adjustment of ST depression for identification of significant coronary artery disease in clinical practice. Int J Cardiol 127: 390–392.
19. IPAQ - International Physical Activity Questionnaires, Available: http://www.ipaq.ki.se/ipaq.htm.
20. Palazoglu M, Fiehn O (2009) Metabolite identification in blood plasma using GC/MS and the agilent fiehn GC/MS metabolomics RTL Library. Agilent Application Note.
21. Albers MJ, Butler TN, Rahwa I, Bao N, Keshari KR, et al. (2009) Evaluation of the ERETIC Method as an improved quantitative reference for H-1 HR-MAS spectroscopy of prostate tissue. Magnetic Resonance in Medicine 61, 525–532.
22. Storey JD, Tibshirani R (2003). Statistical significance for genomewide studies. Proc Nat Acad Sci USA 100: 9440–9445.
23. Sahlin K (1990) Muscle carnitine metabolism during incremental dynamic exercise in humans. Acta Physiol Scandi, 138, 259–262.
24. Bowtell JL, Marwood S, Bruce M, Constantin-Teodosiu D, Greenhaff PL (2007) Tricarboxylic acid cycle intermediate pool size: functional importance for oxidative metabolism in exercising human skeletal muscle. Sports Medicine 37: 1071–1089.
25. Gibala MJ, MacLean DA, Graham TE, Saltin B (1998) Tricarboxylic acid cycle intermediate pool size and estimated cycle flux in human muscle during exercise. Am J of Physiol - End And Metab 275: E235–E242.
26. Chatzinikolaou A, Fatouros I, Petridou A, Jamurtas A, Avloniti A, et al. (2008) Adipose tissue lipolysis is upregulated in lean and obese men during acute resistance exercise. Diabetes Care 31: 1397–1399.
27. Jacob S, Hauer B, Becker R, Artzner S, Grauer P, et al. (1999) Lipolysis in skeletal muscle is rapidly regulated by low physiological doses of insulin. Diabetologia 42: 1171–1174.
28. Stumvoll M, Jacob S, Wahl HG, Hauer B, Löblein K, et al. (2000) Suppression of systemic, intramuscular, and subcutaneous adipose tissue lipolysis by insulin in humans. J Clin Endocrinol Metab 85: 3740–3745.
29. Henriksson J (1991) Effect of exercise on amino acid concentrations in skeletal muscle and plasma. J Exp Biol. 1991 Oct;160: 149–65.
30. Gelfand RA, Barrett EJ (1987) Effect of physiologic hyperinsulinemia on skeletal muscle protein synthesis and breakdown in man. J Clin Invest 80: 1–6.
31. Binko G, Williams BD, Fleming RY, Wolfe RR (1999) Insulin action on muscle protein kinetics and amino acid transport during recovery after resistance exercise. Diabetes 48: 949–957.
32. Lepore M, Pampanelli S, Fanelli C, Porcellati F, Bartocci L, et al. (2000) Pharmacokinetics and pharmacodynamics of subcutaneous injection of long-acting human insulin analog glargine, NPH insulin, and ultralente human insulin and continuous subcutaneous infusion of insulin lispro. Diabetes 49: 2142–2148.
33. Davison GW, George L, Jackson SK, Young IS, Davies B, et al. (2002) Exercise, free radicals, and lipid peroxidation in type 1 diabetes mellitus. Free Radic Biol Med 33: 1543–1551.
34. Yki-Järvinen H, Mott D, Young AA, Stone K, Bogardus C (1987) Regulation of glycogen synthase and phosphorylase activities by glucose and insulin in human skeletal muscle. J Clin Invest 80: 95–100.
35. Yki-Järvinen H, Helve E, Koivisto VA (1987) Hyperglycemia decreases glucose uptake in type 1 diabetes. Diabetes: 36: 892–896.
36. Schauer IE, Snell-Bergeon JK, Bergman BC, Maahs DM, Kretowski A, et al. (2011) Insulin resistance, defective insulin-mediated fatty acid suppression, and coronary artery calcification in subjects with and without type 1 diabetes: The CACTI study. Diabetes 60: 306–314.
37. Bergman BC, Howard D, Schauer IE, Maahs DM, Snell-Bergeon JK, et al. (2012) Features of Hepatic and Skeletal Muscle Insulin Resistance Unique to Type 1 Diabetes. J Clin Endocrinol Metab. 2012. doi: 10.1210/jc.2011–3172.
38. Nadeau KJ, Regensteiner JG, Bauer TA, Brown MS, Dorosz JL, et al. (2010) Insulin resistance in adolescents with type 1 diabetes and its relationship to cardiovascular function. J Clin Endocrinol Metab, 95: 513–521.
39. Perseghin G, Lattuada G, Danna M, Sereni LP, Maffi P, et al. (2003) Insulin resistance, intramyocellular lipid content, and plasma adiponectin in patients with type 1 diabetes. Am J Physiol Endocrinol Metab 285: E1174–1181.
40. Levin K, Dua Schroeder H, Alford FP, Beck-Nielsen H (2001) Morphometric documentation of abnormal intramyocellular fat storage and reduced glycogen in obese patients with Type II diabetes. Diabetologia, 44: 824–833.
41. Kelley DE, Mandarino LJ (2000) Fuel selection in human skeletal muscle in insulin resistance: a reexamination. Diabetes 49: 677–683.
42. Galgani JE, Moro C, Ravussin E (2008) Metabolic flexibility and insulin resistance. Am J Physiol Endocrinol Metab 295: E1009–1017.
43. Mogensen M, Sahlin K, Fernström M, Glintborg D, Vind BF, et al. (2007) Mitochondrial respiration is decreased in skeletal muscle of patients with type 2 diabetes. Diabetes 56: 1592–1599.
44. Befroy DE, Petersen KF, Dufour S, Mason GF, de Graaf RA, et al. (2007) Impaired mitochondrial substrate oxidation in muscle of insulin-resistant offspring of type 2 diabetic patients. Diabetes 56: 1376–1381.
45. Item F, Heinzer-Schweize S, Wyss M, Fontana P, Lehmann R, et al. (2011) Mitochondrial capacity is affected by glycemic status in young untrained women with type 1 diabetes but is not impaired relative to healthy untrained women Am J Physiol Regul Integr Comp Physiol 301: R660–R666.
46. Morino K, Petersen KF, Shulman GI (2006) Molecular mechanisms of insulin resistance in humans and their potential links with mitochondrial dysfunction. Diabetes 55 (Suppl 2): S9–S15.
47. Farhy LS, Chan A, Breton MD, Anderson SM, Kovatchev BP, et al. (2012) Association of fasted hyperglucagonemia with impaired glucagon counter-regulation in type 1 diabetes. Front Physiol. 3:40, doi: 10.3389/fphys.2012.00040.
48. Divertie GD, Jensen MD, Cryer PE, Miles JM (1997) Lipolytic responsiveness to epinephrine in nondiabetic and diabetic humans. Am J Physiol, 272: E1130–1135.
49. Komatsu WR, Gabbay MA, Castro ML, Saraiva GL, Chacra AR, et al. (2005) Aerobic exercise capacity in normal adolescents and those with type 1 diabetes mellitus. Pediatr Diabetes 6: 145–149.
50. Gusso S, Hofman P, Lalande S, Cutfield W, Robinson E, et al. (2008) Impaired stroke volume and aerobic capacity in female adolescents with type 1 and type 2 diabetes mellitus. Diabetologia 51: 1317–1320.
51. Vervoort G, Wetzels JF, Lutterman JA, van Deuren BG, Berden JH, et al. (1999) Elevated skeletal muscle blood flow in noncomplicated type 1 diabetes mellitus: role of nitric oxide and sympathetic tone. Hypertension 34: 1080–1085.